

**KLASIFIKASI TUMOR PAYUDARA PADA CITRA  
MAMMOGRAFI DENGAN METODE SUPPORT VECTOR  
MACHINE (SVM) BERBASIS CIRI TEKSTUR STATISTIK  
MENGUNAKAN SCILAB**

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh  
Gelara Sarjana Fisika  
dalam Ilmu Fisika



Oleh : **UHTY MAESYAROH**  
NIM : 1608026010

FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI WALISONGO  
SEMARANG  
**2022**

## PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama : Uhty Maesyaroh

NIM : 1608026010

Jurusan / Program Studi : Fisika / Fisika

menyatakan bahwa skripsi yang berjudul :

**KLASIFIKASI TUMOR PAYUDARA PADA CITRA MAMMOGRAFI DENGAN METODE SUPPORT VECTOR MACHINE (SVM) BERBASIS CIRI TEKSTUR STATISTIK MENGGUNAKAN SCILAB**

secara keseluruhan adalah hasil penelitian / karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 28 Juni 2022

Pembuat pernyataan



Uhty Maesyaroh

1608026010



KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI WALISONGO  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Prof. Dr. Hamka (Kampus II) Ngaliyan Semarang  
Telp. 024-7601295 Fax. 7615387

### PENGESAHAN

Naskah skripsi berikut :

Judul : **KLASIFIKASI TUMOR PAYUDARA PADA CITRA MAMMOGRAFI DENGAN METODE SUPPORT VECTOR MACHINE (SVM) BERBASIS CIRI TEKSTUR STATISTIK MENGGUNAKAN SCILAB.**

Penulis : **UHTY MAESYAROH**  
NIM : **1608026010**  
Jurusan : **Fisika**

Telah diujikan dalam sidang *tugas akhir* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Fisika.

Semarang, 28 Juni 2022

### DEWAN PENGUJI

Penguji I,

**Heni Sumarti, M.Si.**

NIP : 19871011 201903 2 009

Penguji II,

**Sheila Rully Anggita, M.Si**

NIP : 19900505 201903 2 017

Penguji III,

**Istikomah, M.Sc.**

NIP : 19901126 201903 2 001

Penguji IV,

**Hartono, M.Sc.**

NIP : 19900924 201903 1 000

Pembimbing I

**Muhammad Ardhi K, M. Sc**

NIP: 19821009 201101 1 010

Pembimbing II

**Heni Sumarti, M. Si.**

NIP: 19871011 201903 2 009



NOTA DINAS

Semarang, 28 Juni 2022

Yth. Ketua Program Studi Fisika  
Fakultas Sains dan Teknologi  
UIN Walisongo Semarang

*Assalamu 'alaikum warahmatullahi wabarakatuh*

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul : KLASIFIKASI TUMOR PAYUDARA PADA CITRA MAMMOGRAFI  
DENGAN METODE SUPPORT VECTOR MACHINE (SVM)  
BERBASIS CIRI TEKSTUR STATISTIK MENGGUNAKAN SCILAB  
Nama : Uhty Maesyaroh  
Nim : 1608026010  
Jurusan : Fisika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

*Wassalamu 'alaikum warahmatullahi wabarakatuh*

Pembimbing II,



**Heni Sumarti, M.Si.**

NIP : 198710112019032009

NOTA DINAS

Semarang, 28 Juni 2022

Yth. Ketua Program Studi Fisika  
Fakultas Sains dan Teknologi  
UIN Walisongo Semarang

*Assalamu 'alaikum warahmatullahi wabarakatuh*

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul : KLASIFIKASI TUMOR PAYUDARA PADA CITRA MAMMOGRAFI  
DENGAN METODE SUPPORT VECTOR MACHINE (SVM)  
BERBASIS CIRI TEKSTUR STATISTIK MENGGUNAKAN SCILAB  
Nama : Uhty Maesyaroh  
Nim : 1608026010  
Jurusan : Fisika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

*Wassalamu 'alaikum warahmatullahi wabarakatuh*

Pembimbing I,



**Muhammad Ardhi Khalif, M.Sc.**

NIP : 198210092011011010

## ABSTRAK

Mammografi salah satu teknik pemeriksaan dini untuk mendeteksi kelainan berupa tumor pada payudara. Mammografi menghasilkan citra payudara yang masih memiliki kekurangan berupa *noise*. Sistem *Computer Aided Diagnosis* (CAD) merupakan suatu sistem komputerisasi menggunakan algoritma untuk pengolahan citra. Penelitian ini bertujuan untuk melakukan klasifikasi tumor payudara berbasis CAD sebagai dukungan analisis diagnostic dokter atau ahli radiologi untuk menentukan jenis tumor payudara. Sistem CAD pada penelitian ini menggunakan metode *cropping* dalam penentuan *RoI Region of Interest*, ekualisasi histogram dan *median adaptif filter* dalam peningkatan kualitas citra serta metode ekstraksi fitur tekstur Histogram, *Grey Level Co-occurrence Matrices* (GLCM) dan *Gray Level Run Length Matrices* (GLRLM) sebagai masukan dalam tahap klasifikasi menggunakan *software* Scilab. Tahap klasifikasi menggunakan metode *Support Vector Machine* (SVM) dengan seleksi fitur *Relief*. Hasil akurasi diperoleh sebesar 62,68% dalam membedakan tumor payudara jinak (*benigna*) dan ganas (*malignant*). Nilai akurasi tersebut berdasarkan ROC (*Receiver Operating Characteristic*) untuk penilaian performa metode klasifikasi maka termasuk kategori *poor classification* dalam klasifikasi citra mammogram berdasarkan ciri tekstur statistik.

**Kata kunci** : payudara, citra mammografi, ekstraksi fitur, scilab.

## KATA PENGANTAR

Assalamualaikum Wr. Wb. Alhamdulillah rabbil alamin. Segala puji dan syukur kepada Allah SWT atas rahmat dan barakah-Nya sehingga skripsi dengan judul "KLASIFIKASI TUMOR PAYUDARA PADA CITRA MAMMOGRAFI DENGAN METODE SUPORT VECTOR MACHINE (SVM) BERBASIS CIRI TEKSTUR STATISTIK MENGGUNAKAN SCILAB" dapat penulis selesaikan. Penulisan skripsi ini digunakan sebagai salah satu syarat dalam memperoleh gelar Sarjana Fisika pada Program Studi S1 Fisika UIN Walisongo Semarang.

Proses penyusunan skripsi ini tidak lepas dari doa, bantuan, bimbingan, motivasi dan peran dari banyak pihak. Sehingga penulis mengucapkan terimakasih kepada :

1. Prof. Dr. Imam Taufiq, M.Ag selaku rektor UIN Walisongo Semarang.
2. Dr. Ismail, M.Ag selaku Dekan Fakultas Sains dan Teknologi UIN Walisongo Semarang.
3. Agus Sudarmanto, M.Si selaku Ketua Jurusan Fisika Fakultas Sains dan Teknologi UIN Walisongo.
4. M. Ardhi Khalif, M.Sc dan Heni Sumarti, M.Si selaku dosen pembimbing yang telah dengan penuh kesabaran dan ketulusan dalam memberikan ilmu, bimbingan terbaik serta motivasi kepada penulis, sehingga penelitian dan penyusunan skripsi ini dapat diselesaikan.
5. Segenap Bapak/Ibu dosen dan seluruh Staf Jurusan Fisika Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang yang telah memberikan dan membekali ba-

nyak ilmu baik ilmu agama, pengetahuan maupun moral selama masa perkuliahan.

6. Lasiman dan Ngaisah selaku orang tua penulis serta Yusman dan Suryati selaku mertua penulis terima kasih banyak atas segala do'a, nasehat, semangat, kasih sayang serta motivasi yang tidak dapat terbayarkan oleh apapun. Tak lupa juga Cahyani Ismi Azizah, adikku yang telah memberikan dukungan dan semangat kepada penulis.
7. Kepada suami, Wahyu Andrian S dan anakku Bilal & Bilsha tercinta, terima kasih atas segala dukungan dan pengertiannya selama mamah (penulis) kuliah hingga menyelesaikan penyusunan skripsi ini.
8. Nisrina Nur Aziza, Luthfiyatus Saniyah, dan Nurul Inayah selaku sahabat seperjuangan yang telah memberikan bantuan, dukungan, dan semangat kepada penulis selama proses penyusunan skripsi.
9. Teman-teman Fisika 2016 yang telah menemani penulis selama menempuh pendidikan di Fakultas Sains dan Teknologi UIN Walisongo.
10. Semua Pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya skripsi ini.

Semoga kebaikan semuanya menjadi amal ibadah yang diterima dan mendapat pahala yang berlimpah dari Allah SWT. Aamiin. Atas segala kekurangan dan kelemahan dalam skripsi ini penulis

mengharapkan saran dan kritik yang membangun dari semua pihak. Semoga karya tulis yang sederhana ini dapat menjadi bacaan yang bermanfaat dan dapat dikembangkan bagi peneliti-peneliti selanjutnya. *Aamiin yaa Rabbal aalamiin.*

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	<b>i</b>
<b>PERNYATAAN KEASLIAN</b> .....	<b>iii</b>
<b>PENGESAHAN</b> .....	<b>vi</b>
<b>NOTA PEMBIMBING I</b> .....	<b>vii</b>
<b>NOTA PEMBIMBING II</b> .....	<b>ix</b>
<b>KATA PENGANTAR</b> .....	<b>xv</b>
<b>DAFTAR ISI</b> .....	<b>xvii</b>
<b>DAFTAR TABEL</b> .....	<b>xix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xxi</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xxiii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
A. Latar Belakang Masalah .....	1
B. Rumusan Masalah .....	5
C. Tujuan Penelitian .....	5
D. Manfaat Penelitian .....	5
E. Batasan Masalah .....	6
<b>BAB II LANDASAN PUSTAKA</b> .....	<b>7</b>
A. Kajian Teori .....	7
1. Payudara .....	8
2. Tumor Payudara .....	10
3. Mammografi .....	11
4. <i>Computer Aided Diagnosis (CAD)</i> .....	17
5. Dasar Pengolahan Citra .....	19
6. Ekstraksi Ciri Tekstur .....	21
7. Seleksi Fitur <i>Relief</i> .....	39
8. <i>SVM Support Vector Machine</i> .....	39
9. <i>K-Fold cross validation</i> .....	41
10. <i>Confusion Matrix</i> .....	43
B. Kajian Pustaka .....	44
<b>BAB III METODE PENELITIAN</b> .....	<b>47</b>
A. Waktu dan Tempat Penelitian .....	47

B.	Alat Penelitian .....	47
C.	Bahan Penelitian .....	47
D.	Pengolahan Data .....	48
E.	Analisis Data .....	51
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>53</b>
<b>BAB V PENUTUP .....</b>		<b>67</b>
A.	Kesimpulan .....	67
B.	Saran .....	68
<b>DAFTAR PUSTAKA .....</b>		<b>69</b>
<b>Lampiran-lampiran .....</b>		<b>75</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Judul</b>	<b>Halaman</b>
Tabel 4.1	Citra Hasil Ekualisasi Histogram dan Median Adaptif Filter	55
Tabel 4.2	Rata-rata Hasil Ekstraksi Fitur Tekstur Histogram	56
Tabel 4.3	Rata-rata Hasil Ekstraksi Fitur Tekstur GLCM	57
Tabel 4.4	Rata-rata Hasil Ekstraksi Fitur Tekstur GLRLM	58
Tabel 4.5	Hasil seleksi fitur tekstur	59
Tabel 4.6	Hasil seleksi fitur tekstur	62

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

*Kelenjar mama* atau payudara (buah data) merupakan salah satu organ reproduksi wanita. Pada payudara banyak terdapat saluran limfe (getah bening) yang menjadi saluran penyebaran infeksi. Infeksi pada payudara biasanya dapat menjadi tumor jinak atau ganas. Tumor ganas atau biasa disebut kanker agak lebih sering terjadi di dalam payudara (Evelyn, 2011).

Menurut Kementrian Kesehatan (2019), sebanyak 19,18% penduduk Indonesia menderita kanker payudara dengan rata-rata kematian 17 per 100.000 penduduk. Penanganan yang tepat pada penderita kanker payudara yang melakukan pemeriksaan dini meningkatkan angka kesembuhan sebesar 80-90% serta menurunkan angka kematian sebesar 30%. Pemeriksaan dini pada payudara dapat dilakukan dengan beberapa teknik pemeriksaan seperti ultrasonografi (USG), mammografi dan *Magnetic Resonance Imaging*(MRI).

USG menggunakan gelombang suara berfrekuensi tinggi untuk menghasilkan citra payudara. USG merupakan metode yang relatif murah serta mudah diakses terutama pada daerah yang tidak memadai untuk dilakukan pemeriksaan lain pada pemeriksaan dini payudara. USG payudara mampu membedakan area normal, area padat dan area cairan yang merupakan gambaran dari kista. Namun, USG tidak spesifik dalam membedakan tumor jinak dan tumor ganas karena kurang mampu mendeteksi mikrokalsifikasi yang merupakan salah satu indikasi tumor ganas. Sehingga, USG disa-

rakan dilakukan bersamaan dengan mammografi (Gunawan, 2021).

Mammografi merupakan pemeriksaan payudara dengan menggunakan sinar-x dosis rendah untuk menghasilkan citra payudara. Mammografi mampu menunjukkan mikrokalsifikasi dengan ukuran lebih kecil dari 100  $\mu\text{m}$ , serta sering menunjukkan adanya indikasi tumor bahkan 1-2 tahun sebelum ditemukan dengan pemeriksaan payudara sendiri. Mammografi mampu memberikan resolusi yang sangat baik untuk identifikasi perbedaan antar jaringan lunak, namun kurang sensitif apabila payudara memiliki komposisi berlemak (Gunawan, 2021).

MRI menggunakan teknologi magnet dan gelombang radio untuk menghasilkan citra payudara. MRI menghasilkan gambar berbentuk 2D dan 3D sehingga dalam memvisualisasi kelainan payudara sensitivitasnya hampir 100%. Namun MRI memiliki kelemahan utama yaitu biaya, dimana biaya satu kali MRI diperkirakan setara dengan 10x pemeriksaan mammografi (Gunawan, 2021).

Secara umum, mammografi merupakan pilihan skrining tumor payudara yang utama dan terjangkau. Bahkan menurut Pratama (2014) tingkat akurasi pemeriksaan mamografi dalam mendeteksi keberadaan tumor mencapai 80-90% dari semua kasus tumor payudara. Analisis citra mammogram masih sering dilakukan dengan cara manual, yaitu analisis dari dokter dan radiologis. Liantoni (2020) mengatakan sekarang ini sistem *Computer Aided Diagnosis*(CAD) dapat menjadi salah satu teknik pembelajaran mesin untuk mendiagnosa adanya kelainan pada citra mammogram. Sistem CAD dioperasikan dengan menggunakan algoritma pengolahan citra dan pengenalan pola seperti peningkatan kualitas citra dan penapisan, penggalan dan penyeleksian ciri citra, klasifikasi dan evaluasi hasil (Santoso,2014).

Proses penggalian dan penyeleksian ciri pada citra mammogram terdapat tiga jenis yaitu ciri tingkat keabuan, ciri bentuk dan ciri tekstur (Haralick, 1973). Analisis tekstur dengan metode statistik diklasifikasikan menjadi orde pertama yaitu histogram, orde kedua yaitu *Gray Level Co-Occurrence Matrix* (GLCM) dan orde tingkat tinggi yaitu *Gray Level Run Length Matrix* (GLRLM) (Akdemir, 2018). Ciri yang diperoleh dari analisis tekstur dapat digunakan untuk mengklasifikasi atau mengelompokkan *pixel* berdasarkan ciri-ciri tekstur ke dalam kelas-kelas tertentu. Metode klasifikasi yang banyak digunakan adalah dengan kecerdasan buatan atau *machine learning* salah satunya adalah *Support Vector Machine* (SVM). Pada prosesnya SVM menggunakan sesuatu yang dikenal sebagai *kerneling* untuk memetakan fitur data ke spasi dimensi yang lebih tinggi agar *hyperlane* dapat dibentuk untuk mengklasifikasikannya. *Hyperlane* adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas klasifikasi (Panesar, 2021).

Beberapa penelitian pernah dilakukan pada citra medis untuk proses klasifikasi menggunakan SVM. Kadek dan I Made melakukan klasifikasi citra mammogram dengan mengambil 4 nilai ciri GLCM untuk diklasifikasikan dengan SVM dan didapat akurasi sebesar 80% (Setiawan, 2018). Lussiana melakukan klasifikasi citra mammogram dengan mengambil 3 nilai ciri GLCM untuk diklasifikasikan dengan SVM didapat akurasi sebesar 90% untuk kategori normal dan 87,67% untuk kategori abnormal.

Penelitian tersebut menunjukkan bahwa SVM memiliki tingkat akurasi yang cukup baik dalam proses klasifikasi. Kebanyakan penelitian dengan SVM sebagai metode klasifikasi hanya menggunakan satu dari tiga analisis fitur tekstur statistik yaitu GLCM. Padahal analisis fitur tekstur statistik terdapat tiga jenis yaitu histogram,

GLCM dan GLRLM yang dapat memberikan informasi pada citra serta meningkatkan akurasi pada proses klasifikasi.

Penelitian juga lebih banyak menggunakan software MATLAB. Software Bahasa pemrograman untuk pengolahan citra medis tidak hanya MATLAB saja. Terdapat beberapa pilihan seperti python dan juga scilab. Scilab tersedia gratis dari *Institut Nationale de Recherche en Informatique et en Automatique - INRIA (National Institute for Informatics and Automation Research)* sebagai salah satu bahasa pemrograman numerik dan grafik yang dijalankan menggunakan operasi UNIX, Windows, Linux dan lain-lain (Ermawati, 2020). Scilab dapat dijadikan salah satu pilihan dalam penggunaan bahasa pemrograman untuk pengolahan citra karena sudah dilengkapi fungsi untuk pengolahan citra.

Dari penjelasan tersebut, pada penelitian ini penulis mengaplikasikan SVM sebagai metode pada klasifikasi citra mammogram dengan analisis tekstur statistik yaitu histogram, *Gray Level Co-Occurrence Matrix* (GLCM) dan *Gray Level Run Length Matrices* (GLRLM) untuk mencari nilai-nilai ciri tekstur dengan menggunakan scilab. Proses klasifikasi menggunakan *Support Vector Machine* (SVM) dengan bantuan *machine learning* weka menjadi dua kelas kelompok data yaitu tumor jinak dan tumor ganas (kanker). Hasil penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya serta dapat dijadikan bahan pertimbangan untuk dokter, radiologi, serta fisikawan medis dalam proses analisis area citra yang diduga sebagai tumor.

## **B. Rumusan Masalah**

Rumusan masalah pada penelitian ini yaitu :

1. Bagaimana pengolahan citra mammogram menggunakan metode *Support Vector Machine* (SVM) berdasar ciri tekstur statistik?
2. Berapakah akurasi klasifikasi tumor payudara pada citra *mammogram* menggunakan metode *Support Vector Machine* (SVM) berdasar ciri tekstur statistik?

## **C. Tujuan Penelitian**

Tujuan penelitian ini yaitu :

1. Untuk mengetahui proses pengolahan citra mammogram menggunakan metode *Support Vector Machine* (SVM) berdasar ciri tekstur statistik.
2. Untuk mengetahui tingkat akurasi *Support Vector Machine* (SVM) dalam mengklasifikasi tumor payudara pada citra *mammogram* berbasis ciri tekstur statistik.

## **D. Manfaat Penelitian**

Manfaat yang diharapkan pada penelitian ini antara lain :

1. Bagi Peneliti dan Pembaca

Menambah pengetahuan mengenai pengolahan citra medis baik dari metode pra pengolahan, ekstraksi dan klasifikasi citra medis.

2. Bagi Tenaga Kesehatan Dapat digunakan sebagai opini kedua dalam memberikan diagnosis dan diharapkan membantu dokter, radiologis dan fisikawan medis dalam meningkatkan akurasi diagnosa
3. Bagi Institusi Menambah koleksi bahan Pustaka yang bermanfaat bagi Institusi, khususnya bagi pembaca yang ingin melakukan pengembangan penelitian serupa.

## E. Batasan Masalah

Batasan Masalah pada penelitian ini yaitu :

1. Objek penelitian dibatasi hanya pada citra mammogram payudara wanita dengan arah pengambilan hanya dari samping payudara atau MLO (*Medio Lateral Oblique*) berjumlah 67 citra mammogram ( 42 citra tumor jinak dan 25 citra tumor ganas ). Citra digital mammografi berasal dari *The Mammographic Image Analysis Society* (MIAS) yang diakses dan diunduh melalui laman <http://www.mammoimage.org/databases/>.
2. Penelitian ini mengklasifikasikan citra *mammogram* berdasarkan ciri tekstur menjadi dua kelas kategori, yaitu tumor jinak dan tumor ganas.
3. Pada sistem CAD tahap pra pengolahan sampai ekstraksi ciri tekstur akan dibuat menggunakan software scilab.
4. Proses klasifikasi dengan *Support Vector Machine* SVM menggunakan software WEKA.

## BAB II

### LANDASAN PUSTAKA

#### A. Kajian Teori

Infeksi pada payudara biasanya dapat menjadi tumor jinak atau ganas. Tumor ganas pada payudara lebih sering disebut sebagai kanker payudara. Kanker payudara merupakan kanker dengan jumlah penderita terbanyak. Di Indonesia sendiri jumlah penderitanya menempati peringkat 23 di Asia. Data ini dilansir oleh Kementerian Kesehatan pada tahun 2019. Kanker payudara dapat diderita oleh siapa saja, namun prevalensi tertinggi penderita kanker menurut usia berada pada rentang usia 55-64 tahun. Hingga saat faktor pasti penyebab kanker payudara belum ditemukan. Faktor usia dan riwayat keluarga diduga menjadi salah satu penyebabnya (Sari, 2016). Meski faktor penyebabnya belum diketahui secara pasti, namun kita perlu melakukan pencegahan secara dini oleh diri masing-masing. Salah satunya yaitu melakukan pola hidup sehat untuk menghindari terjangkitnya kanker payudara. Seperti yang dikatakan Rasulullah dalam hadis ;

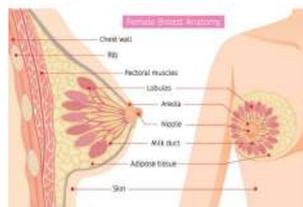
عَنِ ابْنِ عَبَّاسٍ رَضِيَ اللَّهُ عَنْهُمَا قَالَ : قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ: لِرَجُلٍ وَهُوَ يَعِظُهُ: اِغْتَنِمْ خَمْسًا قَبْلَ خَمْسٍ: شَيْبَا بَكَ قَبْلَ هَرَمِكَ: وَصِحَّتَكَ قَبْلَ سَقَمِكَ وَغِنَاءً لَكَ قَبْلَ فَقْرِكَ وَفَرَاغَكَ قَبْلَ شُغْلِكَ وَحَيَاتَكَ قَبْلَ مَوْتِكَ (رواه الحاكم في مستدرک الحاکم)

Artinya: Dari Ibnu Abbas ra: Rasulullah SAW bersabda dan menasehati pada seseorang: "Gunakan yang lima sebelum datang yang lima: masa mudamu sebelum masa tuamu, masa sehatmu sebelum masa sakitmu, masa kayamu sebelum masa miskinmu, masa

lapangmu sebelum masa sibukmu dan masa hidupmu sebelum masa matimu” ( HR. Al Hakim dalam kitab Al Mustadrok). Al Hakim mengatakan bahwa hadist ini shahih sesuai syarat Bukhari Muslim namun keduanya tidak mengeluarkannya). Seperti yang dikatakan dalam hadist diatas, bahwa mencegah lebih baik daripada mengobati. Adanya pencegahan dini seperti pemeriksaan mamografi, diharapkan mampu mengurangi resiko terjangkitnya kanker payudara dengan kasus yang parah sehingga dapat ditangani dengan cepat.

## 1. Payudara

*Kelenjar mama* atau lebih dikenal dengan payudara adalah salah satu organ reproduksi wanita yang mengeluarkan air susu. Letak payudara berada di dalam fasia superfisial di daerah pectoral antara sternum dan aksila dan melebar dari kira-kira iga kedua atau ketiga sampai iga keenam atau ketujuh. Bentuk buah dada cembung ke depan dengan puting di tengahnya, yang terdiri atas kulit dan jaringan erektil dan berwarna tua. Puting dilingkari daerah berwarna coklat yang disebut *areola* (Evelyn, 2011)

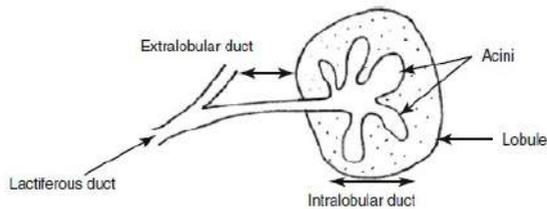


Gambar 2.1. Anatomi Payudara  
(Hogg et al., 2015)

Gambar 2.1 menunjukkan anatomi payudara pada wanita. Secara makroskopis payudara dibagi menjadi dua bagian utama. Ya-

itu jaringan kelenjar yang berkaitan dengan produksi susu dan semua jaringan lain yang menyusun dan menopang payudara. Jaringan tersebut termasuk lemak, fasia (jaringan ikat) dan otot. Jaringan kelenjar terdiri atas 15-20 lobus. Masing-masing lobus terdiri atas 10-100 lobulus yang mengandung banyak sel-sel asinus yang merupakan tempat produksi dan penyimpanan susu selama menyusui (Hogg et al., 2015). Lobulus-lobulus ini disebut sebagai *terminal duct lobular units* ( TDLU ) yang merupakan bagian mikroskopis payudara.

TDLU terdiri dari tiga bagian yaitu sel asini, sebuah saluran intralobular terminalis dan saluran ekstralobularis. Sel asini terdiri dari tiga lapisan yaitu membran basal, lapisan mioepitel dan lapisan epitel. Struktur TDLU diilustrasikan pada gambar 2.2.



Gambar 2.2. Ilustrasi Struktur TDLU  
(Hogg et al., 2015)

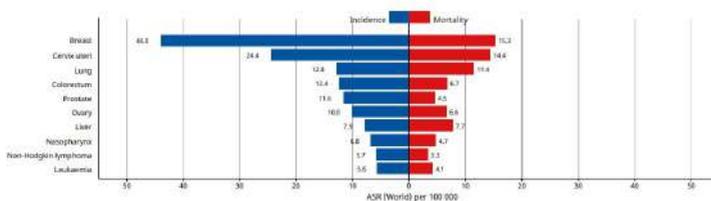
Lapisan epitel biasanya hanya terdapat satu sel tebal, namun dapat menjadi dua atau tiga lapisan sel tebal yang disebut *hyperplasia*. Kondisi *hyperplasia* ini yang merupakan kondisi awal dari tumbuhnya sel tumor. Lebih dari 90% tumor payudara, bermula dari bagian ini. Sedangkan membran basal bertindak sebagai penghalang untuk penyebaran suatu tumor (Hogg et al., 2015).

## 2. Tumor Payudara

Tumor payudara merupakan salah satu kelainan pada payudara. Tumor biasanya ditandai dengan adanya benjolan pada payudara. Benjolan terjadi karena adanya penebalan sel epitel pada bagian kelenjar payudara. Tumor sendiri dikategorikan menjadi dua jenis yaitu tumor jinak dan tumor ganas atau lebih sering disebut sebagai kanker. Keganasan pada jaringan di payudara yang berasal dari lapisan epitel maupun lobulusnya dapat menjadi kanker payudara (Kemenkes RI,2019). Seiring dengan bertambahnya usia, resiko seseorang terkena kanker payudara juga meningkat karena terjadinya mutasi genetik semakin meningkat. Mutasi genetik terjadi akibat adanya kerusakan DNA pada sel normal sehingga menimbulkan pertumbuhan sel yang tidak normal, yang disebut sel kanker. Faktor lain yang dapat memicu tumbuhnya sel kanker antara lain seperti pola makan tidak sehat, penggunaan obat, serta perilaku atau gaya hidup (Hogg et al., 2015).

Di Indonesia kanker payudara merupakan penyebab kematian terbesar akibat kanker. Sebanyak 15,3% kematian dari 44% kasus kanker payudara (Kemenkes RI,2019). Artinya kematian akibat kanker payudara hampir separuh dari jumlah penderita total. Grafik kasus kanker dan tingkat kematian ditampilkan pada gambar 2.3.

Panduan penatalaksanaan kanker payudara yang diterbitkan oleh Kemenkes RI mengatakan bahwa kanker payudara dapat di deteksi dini dengan dua cara pencegahan, yaitu pencegahan primer dan pencegahan sekunder. Pencegahan primer adalah salah satu upaya pencegahan agar terhindar dari kanker payudara dengan cara mengurangi faktor resiko yang diduga erat kaitannya dengan kanker payudara, seperti menerapkan pola hidup sehat. Se-



Gambar 2.3. Data Kasus dan Kematian Akibat Kanker di Indonesia (Globocan,2020)

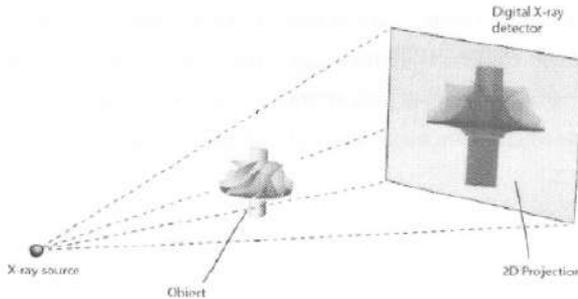
dangkan pencegahan sekunder adalah pencegahan dengan melakukan pemeriksaan abnormalitas pada payudara (skrining) untuk menemukan benjolan yang diindikasikan sebagai tumor payudara. (Kemendes RI,2019)

Beberapa tindakan skrining untuk tumor payudara antara lain : Pemeriksaan Payudara Sendiri (SADARI), Pemeriksaan Payudara Klinis (SADANIS) dan Mammografi Skrining. Skrining dilakukan untuk mendeteksi dini keberadaan tumor payudara sehingga penanganan menjadi lebih efektif untuk menurunkan tingkat kekambuhan dan memperbaiki kualitas hidup seseorang.

### 3. Mammografi

Mammografi merupakan pesawat sinar-x yang digunakan untuk pemeriksaan kelainan pada payudara. Dosis sinar-x yang digunakan umumnya merupakan dosis rendah ( berkisar 0,7 mSv ) karena payudara merupakan jaringan ( *soft tissue* ). Sinar-x memiliki daya tembus yang tinggi sehingga mampu menembus berbagai material. Prinsip kerja sinar-x, Ketika menembus sebuah material, intensitasnya berkurang sebanding dengan jenis material yang dilewatinya. Perbedaan intensitas inilah yang nantinya akan tampak sebagai citra / gambar (Santoso,2014). Prinsip kerja perangkat

sinar-x ditampilkan pada gambar 2.4.

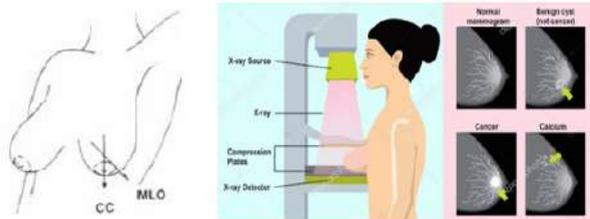


Gambar 2.4. Prinsip Kerja Perangkat Sinar-X  
(Santoso,2014)

Proses sinar-x yang dipakai pada proses mammografi merupakan prinsip sinar-x Bremstrahlung. Bremstrahlung merupakan sinar-x yang terbentuk apabila elektron dengan laju tinggi mengalami suatu percepatan. Tabung vakum penghasil sinar-x berisi logam yang ditembaki elektron sebagai hasil dari pemanasan filamen yang dipercepat karena diberi tegangan listrik berorde  $10^{10}$  volt. Perlambatan terjadi saat elektron berenergi tinggi menabrak logam target sehingga dari permukaan logam terbentuk spektrum sinar-x yang bersifat kontinyu. Proses pembentukan sinar-x inilah kemudian disebut sebagai sinar-x Bremstrahlung (pengereman) (Hogg,2015).

Secara garis besar, pesawat mammografi terdiri dari tiga unit yaitu penghasil sinar-x, unit mekanik dan penangkap citra. Tabung penghasil sinar-x akan memancarkan energi sinar-x melewati payudara kemudian ditangkap oleh kaset film sinar-x yang diserap memiliki variasi berbeda sesuai jenis jaringan. Hasil *exposure* akan tercipta karena variasi penyerapan sehingga memunculkan rincian jaringan payudara dalam bentuk citra/gambar. Citra yang dihasilk-

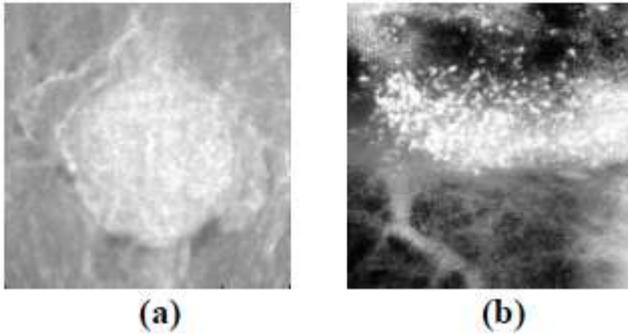
an mammografi disebut sebagai citra *mammogram*. Mammografi mampu memperlihatkan kelainan pada payudara dalam bentuk terkecil hingga kurang dari 5 mm ( stadium nol ) (Santoso,2014).



Gambar 2.5. Titik pandang, proses pengambilan dan citra hasil (Santoso,2014)

Gambar 2.5 merupakan proses pengambilan citra payudara dengan mammografi, citra yang dihasilkan serta pengambilan titik pandang citra. Pengambilan citra payudara menggunakan dua titik pandang pada masing-masing payudara kanan dan kiri. Titik pandang pertama dari samping payudara disebut MLO (*Medio-Lateral Oblique*) dan titik pandang kedua dari atas kebawah payudara disebut CC (*Cranio-Caudal*) (Indrati et al., 2009)

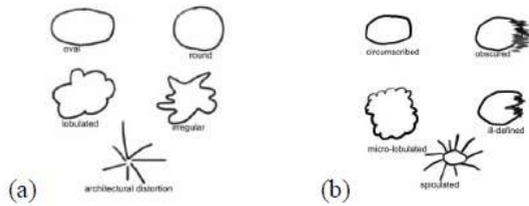
Pada mammogram, warna citra akan nampak hitam, putih dan abu-abu. Warna hitam menunjukkan lemak sedangkan warna putih dan abu-abu menunjukkan kelenjar, jaringan ikat, tumor dan kalsifikasi muneul (Santoso,2014). Kanker payudara dapat dikenali dengan keberadaan lesi massa atau keberadaan mikrokalsifikasi pada citra mammografi. Lesi masa didefinisikan sebagai lesi dengan bentuk, tekstur dan batas area tertentu pada proyeksi citra mammografi (Berment et al., 2014). Sedangkan mikrokalsifikasi merupakan noda atau titik-titik kecil yang nampak pada citra mammografi. Mikrokalsifikasi biasanya terdapat pada area lobula atau ductal payudara (Putra, 2009).



Gambar 2.6. Potongan citra (a) massa (b) mikrokalsifikasi  
(Putra, 2009)

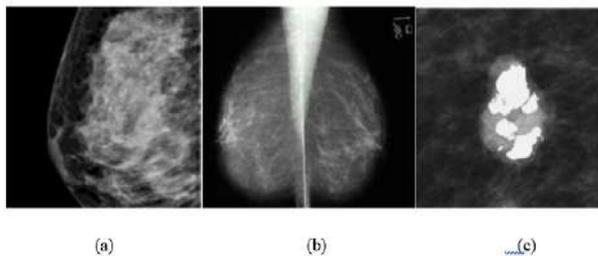
Hasil pemeriksaan mammografi diklasifikasikan menurut panduan *The American College of Radiology* yang dikenal sebagai *Breast Imaging, Reporting & Data System* (BIRADS) (Ermawati, 2020). Menurut sistem BIRADS, lesi masa dibedakan oleh :

1. Bentuk : *round, oval, lobulate* atau *irregular*
2. Batas tepi : (a) *circumscribed* antara luka dan jaringan sekitar terlihat jelas, (b) *microlobulated* berbentuk lingkaran dengan sepanjang tepi yang berombak, (c) *obscured* Sebagian jaringan normal menutupi, (d) *ill-defined* bersifat menyebar dan (d) *spiculated* berupa garis tipis yang menyebar
3. Kepadatan : rendah, lebih padat dari jaringan sekitarnya atau berlemak
4. Berhubungan dengan kondisi lain : mikrokalsifikasi, retraksi kulit, retraksi puting, penebalan kulit.

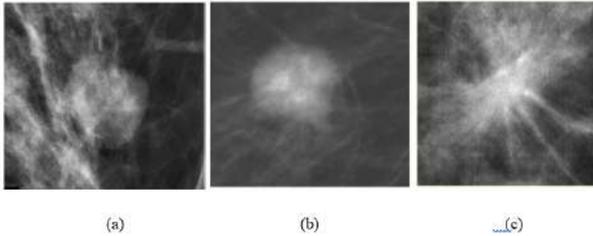


Gambar 2.7. Lesi masa berdasarkan (a) bentuk (b) batas tepi (Indrati et al., 2009)

Lesi masa inilah yang selanjutnya digunakan sebagai dasar untuk menentukan kelainan pada payudara termasuk tumor jinak atau kanker. Benjolan yang berbentuk oval atau round mengidentifikasi tumor jinak, batas tepi yang mempunyai bentuk *ill-defined* dan *spiculated* mempunyai kemungkinan besar tumor ganas. Tingkat kepadatan rendah mengidentifikasi tumor jinak, dan mikrokalifikasi mempunyai kemungkinan tumor ganas. Berikut merupakan klasifikasi citra mamografi berdasarkan BIRADS (Berment et al., 2014)



Gambar 2.8. (a) BIRADS-0 (b) BIRADS-1 (c) BIRADS-2 (Berment et al., 2014)



Gambar 2.9. (a)BIRADS-3 (b)BIRADS-4 (a)BIRADS-5  
(Berment et al., 2014)

Pada gambar 2.8 merupakan klasifikasi BIRADS-0, BIRADS-1 dan BIRADS-2. BIRADS-0 dikategorikan tidak lengkap namun citra memiliki jaringan yang sangat padat sehingga perlu adanya evaluasi pencitraan tambahan. BIRADS-1 dikategorikan negatif, tidak tampak kelainan maupun masa yang terbentuk sehingga tidak perlu adanya tindakan tambahan. BIRADS-2 dikategorikan jinak dengan batas tegas dan terdapat fibriadenoma terkalsifikasi berbentuk karang koral.

Pada gambar 2.9 merupakan kategori BIRADS-3, BIRADS-4 dan BIRADS-5. BIRADS-3 dikategorikan mungkin jinak dengan resiko keganasan lebih rendah dari 2%. Masa berbentuk bulat dengan batas tepi *circumscribed* sehingga perlu ada tindak lanjut dengan jangka pendek (6 bulan). BIRADS-4 dikategorikan curiga abnormal dengan resiko keganasan 2% hingga 94%. Masa dengan bentuk dan batas tepi yang tidak jelas sehingga tindakan biopsi perlu dipertimbangkan. BIRADS-5 dikategorikan sangat ganas dengan resiko keganasan diatas 94%. Pada gambar diatas masa berbentuk *irregular* dengan batas tepi *spiculated*. Untuk kasus dengan kategori ini, tindakan yang tepat harus diambil

BIRADS – 6 termasuk kanker. Disarankan melakukan biopsi.

Temuan ini dikelola sebagai kanker payudara dini (Ermawati, 2020).

#### **4. Computer Aided Diagnosis (CAD)**

Pada tahun 1960 upaya awal komputerisasi citra medis dilakukan, lalu pada tahun 1980 an pengembangan berbagai skema CAD mulai dilakukan. Bahkan saat ini, *Computer Aided Diagnosis* (CAD) telah menjadi prosedur rutin klinis untuk mendeteksi kanker payudara pada mammogram di banyak klinik dan pusat kesehatan. Dengan CAD, ahli radiologi dapat menjadikan sebagai pertimbangan dalam diagnostik. Hal ini karena hasil CAD dapat dimasukkan analisis kuantitatif oleh dokter sebagai diagnostic (Marques, 2018).

*Computer Aided Diagnosis* (CAD) didefinisikan sebagai diagnosis yang dibuat oleh ahli radiologi atau dokter menggunakan output dari skema komputerisasi untuk analisis gambar sebagai bantuan diagnostic. Terdapat dua variasi CAD yaitu ; CADe untuk mendeteksi kelainan dari bagian tertentu citra yang disebut *Region Of Interest* (ROIs) dan CADx untuk klasifikasi ROIs yang terdeteksi ada atau tidak adanya penyakit, seperti kanker. Cara kerja dari sistem CAD menurut Marques dalam bukunya yang berjudul *Medical Image Analysis and Informatics : Computer Aided Diagnosis and Therapy* adalah sebagai berikut :

1. Preprocessing citra untuk analisis lebih lanjut
2. Deteksi dan segmentasi ROIs
3. Ekstraksi citra untuk analisis kuantitatif
4. Seleksi fitur citra
5. Training untuk proses klasifikasi citra, dan

## 6. Klasifikasi pola dan pengambilan keputusan diagnostic

Pengembangan skema CAD banyak dilakukan dengan *software* bahasa pemrograman. Salah satu yang banyak dipakai adalah *software* MATLAB. MATLAB merupakan produk komersil dari perusahaan Mathworks, Inc yang merupakan bahasa pemrograman komputasi numerik untuk berbagai perhitungan teknik dan ilmiah (Hocque, 2005).

*Software* lain juga dapat digunakan untuk pengembangan skema CAD, seperti *software* scilab. Scilab merupakan bahasa pemrograman komputasi numerik dengan algoritma yang mencakup banyak aspek terutama masalah komputasi ilmiah. Sejak 1994, scilab disediakan secara gratis oleh *Institut Nationale de Recherche en Informatique et en Automatique – INRIA (National Institute for Informatics and Automation Research)*. Sistem operasi seperti UNIX, Windows, Linux dan lainnya dapat menjalankan *software* scilab. (Baudin, 2010). Pada awalnya scilab difokuskan pada aljabar linier, namun saat ini fiturnya diperluas mencakup banyak bidang komputasi ilmiah. Salah satunya adalah fitur untuk pengolahan citra, mulai dari peningkatan citra, analisis statistik dan pemodelan. *Software* scilab pada penelitian ini digunakan untuk membuat sistem *preprocessing* sampai dengan proses ekstraksi

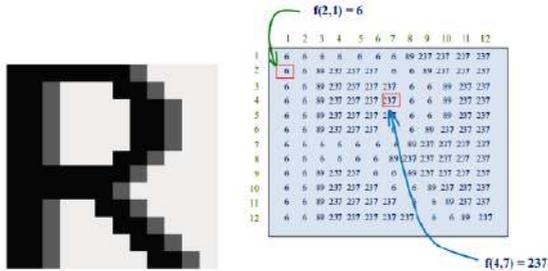
Proses klasifikasi menggunakan bantuan *Machine Learning* Weka. Weka merupakan perangkat lunak yang digunakan untuk pembelajaran dan pelatihan data dengan menggunakan bahasa pemrograman Java. Proses pengumpulan informasi seperti klasterisasi dan klasifikasi dapat dilakukan tanpa harus membangun kode. Hal ini karena weka sudah menyediakan fasilitas tersebut untuk menganalisis data (Frank et al., 2010)

## 5. Dasar Pengolahan Citra

Pengolahan citra adalah istilah umum untuk berbagai teknik yang keberadaannya untuk memanipulasi dan memodifikasi citra dengan berbagai cara. Pengolahan citra merupakan bagian penting yang mendasari bagian aplikasi nyata seperti pengenalan pola dan penginderaan jarak jauh melalui satelit atau pesawat udara. Pada pengenalan pola, pengolahan citra berperan memisahkan objek dari latar belakang secara otomatis yang selanjutnya diproses oleh pengkalsifikasi pola (Kadir & Susanto, 2013).

Citra digital dibentuk oleh kumpulan titik yang dinamakan *pixel* dengan koordinat  $(x,y)$ . koordinat  $x$  menyatakan posisi kolom dan koordinat  $y$  menyatakan posisi baris. Warna, tekstur dan corak tertentu direpresentasikan oleh nilai intensitas setiap *pixel* (Yusu-fiyah, 2016). Pada pemrosesan citra umumnya terdapat tiga jenis citra yang biasa digunakan yaitu citra berwarna, citra berskala keabuan dan citra biner.

Citra berwarna atau citra RGB, merupakan citra yang menyajikan warna dalam bentuk komponen R (merah), G (hijau) dan B (biru). Setiap komponen warna menggunakan delapan bit atau sekitar antara 0 sampai dengan 255. Kemungkinan warna yang disajikan mencapai  $255 \times 255 \times 255$  atau 16.581.375 warna. Citra berskala keabuan menangani gradasi warna hitam dan putih sehingga menghasilkan efek warna abu-abu. Warna dinyatakan dengan intensitas berskala 0 sampai dengan 255 dengan 0 menyatakan hitam dan 255 menyatakan putih. Citra biner merupakan citra dengan tiap *pixel* yang hanya dinyatakan dengan sebuah nilai dari dua kemungkinan yaitu 0 dan 1. Nilai 0 menyatakan warna hitam dan nilai 1 menyatakan warna putih (Kadir & Susanto, 2013).



Gambar 2.10. Contoh notasi intensitas *pixel* pada citra 12x12 (Yusufiyah, 2016)

Gambar 2.10 merupakan contoh notasi intensitas *pixel* untuk citra huruf R dengan ukuran 12x12. Sesuai gambar diatas nilai intensitas  $f(4,7)$  sebesar 237 dan  $f(2,1)$  sebesar 6. Nilai intensitas yang semakin kecil mendekati nol maka citra akan semakin gelap, sedangkan nilai intensitas yang semakin besar mendekati 255 maka warna citra semakin mendekati warna putih. Ini karena pada citra berskala abu-abu nilai intensitas dari 0 hingga 255 dengan 0 menyatakan warna hitam dan 255 menyatakan warna putih (Yusufiyah, 2016). Cara kerja CAD yang paling awal yaitu proses *pre-processing*. Dimana pada proses ini dilakukan peningkatan kualitas citra yang merupakan proses penting sebelum masuk ke tahap pengolahan citra berikutnya (ekstraksi ciri dan klasifikasi). Peningkatan kualitas citra secara kuantitatif hasilnya menjadi lebih baik. Pada citra mammogram, citra biasanya memiliki keterbatasan seperti adanya *noise* yang timbul saat skrining atau kontras citra yang rendah (Fajrin, 2016). Untuk memperbaiki kualitas citra pada penelitian ini yaitu dengan ekualisasi histogram dan median adaptif filter. Histogram ekualisasi merupakan proses pemerataan histogram supaya citra memiliki distribusi nilai derajat keabuan yang

rata. Histogram ekualisasi bekerja untuk mendapatkan aras keabuan yang lebih luas pada daerah yang memiliki banyak *pixel* dan mempersempit aras keabuan pada daerah dengan *pixel* yang sedikit (Riyanto dkk, 2020). Setelah dilakukan histogram ekualisasi citra, maka selanjutnya citra dilakukan proses *filtering*. Terdapat dua jenis filter yaitu filter linear dan filter non-linear. Filter linear contohnya yaitu *mean*, *Laplacian* dan *Laplacian of Gaussian*. Sedangkan filter non-linear contohnya minimum, median, median adaptif dan maksimum filter. Filter non-linear yang cukup populer salah satunya adalah median adaptif filter. Median adaptif filter merupakan modifikasi dari median filter. Apabila median filter bekerja dengan mengambil nilai tengah tingkat *brightness* dari *pixel* sekitarnya untuk mengganti nilai asli intensitas *pixel* maka median adaptif filter bekerja dengan menghitung nilai maksimum, minimum dan nilai tengah tingkat *brightness* untuk menetapkan nilai baru pada *pixel* tertentu (Yusufiyah, 2016).

Cara kerja CAD yang selanjutnya yaitu segmentasi citra. Segmentasi citra, memisahkan citra ke beberapa bagian yang kemudian bagian-bagian tersebut diberi nama sebagai segmen. Objek dan latar belakang (*background*) dapat menjadi segmen. Proses segmentasi penting apabila informasi yang digali pada citra merupakan informasi bentuk kanker. Pada penelitian ini, informasi yang digali merupakan ciri-ciri tekstur sehingga pada penelitian ini tidak menggunakan proses segmentasi citra.

## 6. Ekstraksi Ciri Tekstur

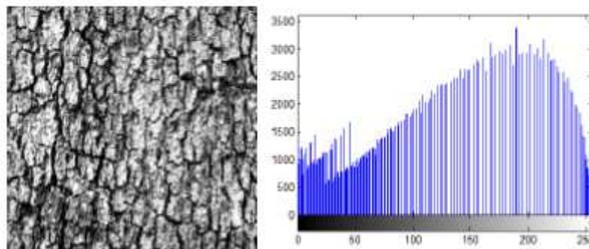
Secara umum tekstur didefinisikan sebagai pengulangan teratur dari sebuah elemen atau pola pada permukaan gambar. Pola-pola tersebut dapat berupa titik, garis atau luasan dengan interval

tertentu yang muncul berulang sehingga karakteristik pengulangannya dapat ditentukan (Listyaningrum,2010). Ciri tekstur berkaitan dengan tingkat kekasaran (*roughness*), granularitas (*granularity*), dan keteraturan (*regularity*) susunan struktural pixel (Haralick, 1973). Ekstraksi ciri tekstur banyak menggunakan proses statistik untuk menggali informasi dari citra. Metode statistik sendiri dibagi menjadi tiga orde. Statistik orde pertama yaitu histogram, statistik orde kedua *Gray Level Co-occurrence Matrice* (GLCM) dan statistik orde tinggi *Gray Level Run Length Matrice* (GLRLM) (Nugroho, 2015).

Histogram akan menganalisa perbedaan nilai frekuensi yang muncul. GLCM akan menganalisa dari berbagai sudut untuk tingkat keseragaman antar *pixel*. Sedangkan GLRLM akan menganalisa panjang *pixel* di seluruh resolusi citra (Ermawati, 2020). Diharapkan dengan perpaduan ciri statistik orde satu, orde dua dan orde tinggi mampu mengenali pola tekstur pada citra secara lebih baik.

#### a. Histogram

Histogram merupakan metode statistik untuk mendapatkan tekstur yang paling sederhana. Pada histogram, nilai intensitas *pixel* citra digambarkan dalam bentuk diagram frekuensi. Intensitas yang sangat banyak pada *pixel* akan dinyatakan dengan nilai yang besar. Histogram dalam pengolahan citra digunakan untuk meningkatkan kontras dan kecerahan citra. Pola tekstur citra histogram terdapat pada gambar 2.11.



Gambar 2.11. Histogram pada citra  
(Nugroho, 2015)

Histogram menurunkan ciri tekstur berdasar perhitungan nilai *pixel* citra asli dengan tidak memperhatikan hubungan ketetanggan *pixel* (Ermawati, 2020). Ciri-ciri yang diturunkan dinamakan ciri statistik orde satu. Nilai ciri-ciri tersebut yaitu:

1. Mean

$$m = \sum_{i=0}^{L-1} i.p(i) \quad (2.1)$$

Nilai rata-rata kecerahan objek akan dihasilkan dari rumus diatas. Aras keabuan pada citra  $f$  dinyatakan dengan  $i$ , probabilitas kemunculan  $i$  dinyatakan dengan  $p(i)$  dan aras keabuan dengan nilai tertinggi dinyatakan dengan  $L$ .

2. Modus

Pada keseluruhan *pixel* nilai aras keabuan yang paling banyak muncul dinyatakan dengan nilai modus. Pada histogram, nilai modulusnya merupakan aras keabuan pada puncak tertinggi kurva. Sebagai contoh, pada gambar 2.13 nilai modulusnya berkisar antara 150-160.

## 3. Energi

$$Energi = \sum_{i=0}^{L-1} [p(i)]^2 \quad (2.2)$$

Energi atau keseragaman menyatakan distribusi intensitas *pixel* terhadap jangkauan aras keabuan. Energi maksimum akan bernilai 1 apabila citra seragam dengan satu nilai aras keabuan.

## 4. Entropi

$$Entropi = - \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (2.3)$$

Tingkat kekompleksan suatu citra direpresentasikan dengan nilai entropi. Makin besar nilainya, artinya sebuah citra semakin kompleks.

5. *Skewness*

$$Skewness = \sum_{i=1}^{L-1} (i - m)^3 p(i) \quad (2.4)$$

Ukuran ketidaksimetrisan terhadap rerata intensitas pada suatu citra dinyatakan dengan *skewness*. Nilainya akan (+) apabila kurva histogram condong ke arah kanan dari nilai rerata intensitas dan nilainya akan (-) apabila sebaliknya. Nilai  $m$  pada persamaan merupakan nilai *mean* yang didapat dari persamaan satu.

## 6. Kurtosis

$$Kurtosis = \sum_{i=1}^{L-1} (i - m)^4 p(i) - 3 \quad (2.5)$$

Keruncingan suatu kurva histogram akan dinyatakan dengan

kurtosis. Nilai kurtosis 0 untuk kurva histogram dengan distribusi normal. Kurtosis dengan nilai  $(-)$  artinya kurva cenderung runcing dan nilai  $(+)$  artinya kurva cenderung melebar.

#### 7. Deviasi Standar

$$\sigma = \sqrt{\sum_{i=1}^{L-1} (i - m)^2 p(i)} \quad (2.6)$$

Tingkat penyebaran data berdasar nilai rata-rata akan dinyatakan dengan standar deviasi.

#### 8. Variance

$$\sigma^2 = \sum_{i=1}^{L-1} (i - m)^2 p(i) \quad (2.7)$$

Ukuran kekontrasan akan diberikan oleh ciri *variance* ini.  $P(i)$  adalah suatu fungsi peluang sehingga  $\sigma^2$  dinamakan dengan *variance*.

#### 9. Smoothness

$$Smoothness = 1 - \frac{1}{1 + \sigma^2} \quad (2.8)$$

Tingkat kehalusan sebuah citra dinyatakan dengan *smoothness*. Citra dengan intensitas kasar memiliki nilai *smoothness* yang rendah.

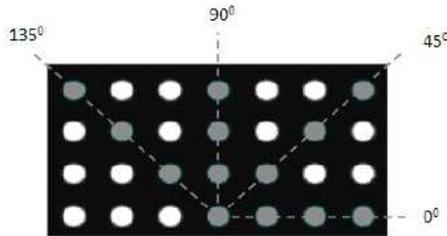
#### b. Grey Level Co-occurrence Matrices ( GLCM )

Ekstraksi ciri dengan metode *Grey Level Co-occurrence Matrices* (GLCM) merupakan perhitungan statistik orde kedua dengan

menggunakan perhitungan tekstur. GLCM pertama kali dipublikasikan pada tahun 1973 oleh Haralick untuk merepresentasikan corak pada tekstur citra. Apabila pada statistik orde pertama menu-runkan ciri tekstur berdasar perhitungan nilai *pixel* citra asli dengan tidak memperhatikan hubungan ketetanggaan *pixel*, maka pada statistik orde kedua hubungan ketetanggaan *pixel* citra diperhitungkan (Haralick, 1973) Misalnya, citra  $f(x,y)$  dengan ukuran  $N_x$  arah horisontal dan  $N_y$  arah vertikal dengan kemungkinan *pixel* hingga  $L$  level dan vektor arah spasial  $\vec{r}$  maka rumus  $GLCM_{\vec{r}}(i, j)$  sebagai berikut :

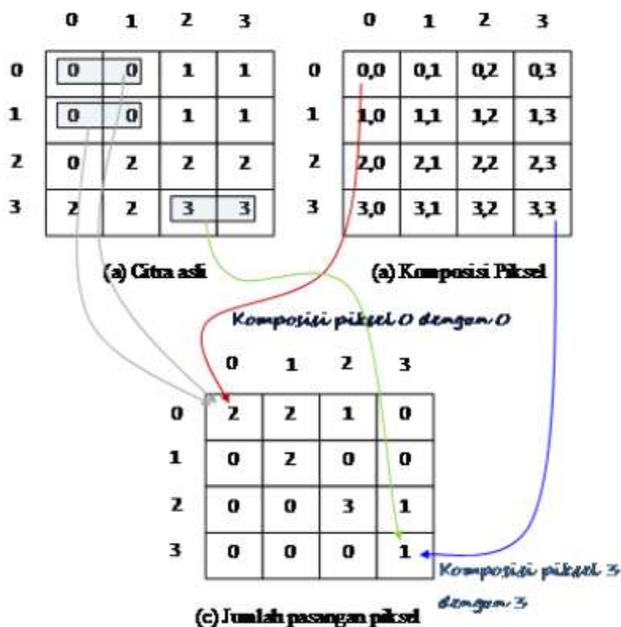
$$GLCM_{\vec{r}}(i, j) = \#\{(x_1, y_1), (x_2, y_2) \in (N_x, N_y) \times (N_x, N_y) | f(x_1, y_1) = i \wedge f(x_2, y_2) \wedge \vec{r} = (x_2 - x_1, y_2 - y_1)\} \quad (2.9)$$

$GLCM_{\vec{r}}(i, j)$  merupakan penjumlahan *pixel*  $j \in 1 \dots L$  yang terjadi pada arah  $\vec{r}$  terhadap *pixel*  $i \in 1 \dots L$ . arah yang dimaksud merupakan arah sudut dengan interval  $45^\circ$  yaitu sudut  $0^\circ, 45^\circ, 90^\circ, 135^\circ$ . Arah-arah sudut pada GLCM akan ditunjukkan pada gambar 2.12.



Gambar 2.12. Arah sudut GLCM  
(Nugroho, 2015)

Untuk penentuan matriks GLCM diberikan contoh dengan sudut  $0^\circ$  dan jarak 1 *pixel* pada gambar 2.13.a. Notasi untuk mempresentasikan hubungan *pixel* dapat ditulis dengan (r,c). sebagai contoh pada gambar 2.13 (b) dan (c) hubungan dua *pixel* berjajar horizontal dengan sudut  $0^\circ$  ditulis dengan notasi (0,0) dengan *pixel* bernilai 0 lalu diikuti *pixel* bernilai 0 disebelah kanan. Begitu pun dengan notasi (3,3) merupakan dua *pixel* berjajar horizontal dengan sudut  $0^\circ$ , dimana *pixel* bernilai 3 diikuti sebelah kanannya *pixel* bernilai 3. Matriks dari hasil perhitungan tersebut dinamakan *matrix framework*, namun masih perlu diolah dengan menambahkan hasil transposnya agar menjadi matriks simetris seperti pada gambar 2.14



Gambar 2.13. Pasangan *pixel* untuk penentuan matriks GLCM  $0^\circ$  (Nugroho, 2015)

$$\begin{array}{ccc}
 \begin{bmatrix} 2 & 2 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} & + & \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & = & \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \\
 \text{Matrik} & & \text{Matrik} & & \text{Matrik} \\
 \text{framework} & & \text{framework} & & \text{framework} \\
 \text{lama} & & \text{baru} & & \\
 & \text{Transpos} & & & 
 \end{array}$$

Gambar 2.14. Pembentukan matriks simetris  
(Nugroho, 2015)

Elemen matriks GLCM masih bergantung pada resolusi citra sehingga perlu dihilangkan dengan cara dinormalisasi. Proses normalisasi akan menghasilkan matriks dengan rentang 0 sampai 1. Normalisasi merupakan perhitungan dengan cara membagi setiap elemen matriks GLCM dengan jumlah total elemen matriks *framework* yang baru. Pada gambar 2.14 jumlah total elemen matriks *framework* yang baru adalah 24. Matriks GLCM yang telah dinormalisasi ditunjukkan pada gambar 2.15.

$$\begin{bmatrix} \frac{4}{24} & \frac{2}{24} & \frac{1}{24} & \frac{0}{24} \\ \frac{2}{24} & \frac{4}{24} & \frac{0}{24} & \frac{0}{24} \\ \frac{1}{24} & \frac{0}{24} & \frac{6}{24} & \frac{1}{24} \\ \frac{0}{24} & \frac{0}{24} & \frac{1}{24} & \frac{2}{24} \end{bmatrix}$$

Gambar 2.15. Pembentukan matriks simetris  
(Nugroho, 2015)

Setelah matriks GLCM ternormalisasi, nilai karakteristik tekstur dari matriks GLCM dapat digali. Penelitian (Ermawati, 2020) menggunakan 21 persamaan untuk menggali nilai-nilai karakteristik tekstur tersebut. Pada penelitian ini, penulis menambahkan dua ciri GLCM yang didapat pada penelitian (Husnul & Yuslena, 2016)

(Verma et al, 2020). Jumlah total ciri yang digunakan oleh penulis yaitu sebanyak 23 ciri. Notasi yang akan digunakan yaitu :

$p(i, j)$  merupakan matrik  $GLCM\vec{r}(i, j)$  dengan ukuran  $i$  baris dan  $j$  kolom yang ternormalisasi.  $p_x(i)$  merupakan matrik baru hasil penjumlahan seluruh elemen baris pada matrik  $p(i, j)$  dapat dituliskan:

$$p_x = \sum_{j=1}^{N_g} p(i, j) \quad (2.10)$$

nilai  $N_g$  sama dengan  $L$  adalah jumlah kuantisasi aras keabuan yaitu 256.

$$p_y = \sum_{j=1}^{N_g} p(i, j) \quad (2.11)$$

$$p_{x+y}(k) = \sum_{i=1} \sum_{j=1+i+j=k} p(i, j) \quad (2.12)$$

dengan  $k = i + j; k = 0, 1, 2...2N_g$

$$p_{x-y}(k) = \sum_{i=1} \sum_{j=1+i+j=k} p(i, j) \quad (2.13)$$

dengan  $k = |i - j|; k = 0, 1, 2...2N_g - 1$

Adapun 24 ciri yang digunakan, dirumuskan sebagai berikut :

#### 1) *Contrast*

$$f_1 = \sum_{i=1} \sum_{j=1} [i - j]^2 p(i, j) \quad (2.14)$$

2) *Angular second moment (ASM) / Energy*

$$f_2 = \sum_i \sum_j \{p(i, j)^2\} \quad (2.15)$$

3) *Autocorrelation (ACorr)*

$$f_3 = \sum_i \sum_j (ij) \cdot p(i, j) \quad (2.16)$$

4) *Dissimilarity (Dissi)*

$$f_4 = \sum_i \sum_j (i - j)p(i, j) \quad (2.17)$$

5) *Inertia*

$$f_5 = \sum_i \sum_j (i - j)^2 p(i, j) \quad (2.18)$$

6) *Joint Average*

$$f_6 = \sum_i \sum_j p(i, j) \quad (2.19)$$

7) *Joint Entropy*

$$f_7 = - \sum_i \sum_j p(i, j) \log p(i, j) \quad (2.20)$$

8) *Sum Of Square / Joint Variance ( $\sigma^2$ )*

$$f_8 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (2.21)$$

dengan  $\mu$  adalah normalisasi dari rerata nilai matriks  $GLCMp(i, j)$

9) *Correlation*

$$f_9 = \frac{\sum_i = 1 \sum_j = 1 (i - \mu_x)(j - \mu_y)p(i, j)}{\sigma^2} \quad (2.22)$$

dengan  $\mu_x, \mu_y$  adalah nilai rerata dari matriks  $p_x, p_y$  sedangkan  $\sigma^2$  merupakan nilai dari *joint variance*

$$\mu_x = \sum_i \sum_j i \cdot p(i, j) \quad (2.23)$$

$$\mu_y = \sum_i \sum_j j \cdot p(i, j) \quad (2.24)$$

10) *Variance*

$$f_{10} = \sum_i \sum_j (i - \mu_x)(j - \mu_y)p(i, j) \quad (2.25)$$

11) *Cluster Shade (Cshade)*

$$f_{11} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j) \quad (2.26)$$

12) *Cluster Prominance (CProm)*

$$f_{12} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j) \quad (2.27)$$

13) *Maximum Probability (Mprob)*

$$f_{13} = \sum_i \sum_j \max p(i, j) \quad (2.28)$$

14) *Sum Average*(SAV)

$$f_{14} = \sum_{i=2}^{2N_g} i \cdot p_{x+y}(i) \quad (2.29)$$

15) *Sum Entropy*(SEntro)

$$f_{15} = - \sum_{i=2}^{2N_g} p_{x+y}(i) \log_2 \{p_{x+y}(i)\} \quad (2.30)$$

16) *Sum Variance*(SVar)

$$f_{16} = \sum_{i=2}^{2N_g} (i - f_{14})^2 p_{x+y}(i) \quad (2.31)$$

17) *Difference Entropy*(DEntro)

$$f_{17} = \sum_{i=0}^{N_g-1} p_{x-y}(i) \log \{p_{x-y}(i)\} \quad (2.32)$$

18) *Difference Variance* (DVar)

$$f_{18} = \text{variandarip}_{x-y} \quad (2.33)$$

19) *Inverse Difference* (INV)

$$f_{19} = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|} \quad (2.34)$$

20) *Invers Difference Normalized (INN)*

$$f_{20} = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|/N_g} \quad (2.35)$$

21) *Invers Different Moment (IDM)*

$$f_{21} = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|^2} \quad (2.36)$$

22) *Invers Difference Moment Normalized (IDMN)*

$$f_{22} = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|^2/N_g^2} \quad (2.37)$$

23) *Information Measures of Correlation 1 (IMCorr 1)*

$$f_{23} = \frac{HXY - HXY1}{\max(HX, HY)} \quad (2.38)$$

dengan

$$HXY = f_6 \quad (2.39)$$

$$HXY1 = -\sum_i \sum_j p(i, j) \log_2 \{p_x(i)p_y(j)\} \quad (2.40)$$

$$HX = \text{Entropydarip}_x \quad (2.41)$$

$$HY = \text{Entropydarip}_y \quad (2.42)$$

24) *Information Measures of Correlation 2 (IMCorr 2)*

$$f_{24} = (1 - \exp[-2(HXY2 - HXY)])^{1/2} \quad (2.43)$$

dengan

$$H_{XY} = - \sum_i \sum_j p_x(i)p_y(j) \log_2\{p_x(i)p_y(j)\} \quad (2.44)$$

Jumlah total keseluruhan ciri statistik orde satu GLCM ada 24. Jika pada statistik orde satu histogram pola intensitas dari sudut arah tidak dihasilkan, maka pada GLCM ini menghasilkan pola intensitas *pixel* dari sudut berturut-turut pada arah 0°, 45°, 90° dan 135°. Diharapkan dengan GLCM dapat membedakan antara dua citra yang memiliki histogram yang sama dengan corak persebaran *pixel* yang berbeda.

### c. *Gray Level Run Length Matrices* (GLRLM)

*Gray Level Run Length Matrices* (GLRLM) merupakan metode untuk merepresentasikan tekstur yang dipublikasi pada tahun 1975 oleh Galloway. Algoritma GLRLM hampir sama dengan GLCM, akan tetapi GLRLM lebih baik dalam memetakan pola-pola tekstur yang memiliki kesamaan panjang piksel. *Gray Level* adalah sebuah nilai aras intensitas citra yang sama secara berturut-turut baik secara vertical, diagonal maupun horizontal. *Run Length* merupakan jumlah *pixel* yang ditempati oleh nilai aras intensitas yang dimaksud. GLCM hanya mampu menganalisis kesamaan tekstur pada *Run Length* ( $d$ ) yang sama. Pola tekstur dengan jarak  $d$  yang berbeda-beda tidak mampu dianalisis pada algoritma GLCM (Nugroho, 2015).

Pada penelitian Nugroho (2015) dijelaskan bahwa statistik orde satu histogram hanya menggunakan satu variable aras intensitas untuk membedakan pola tekstur, statistik orde dua GLCM menggunakan variable intensitas dan sudut untuk membedakan corak tekstur sedangkan statistik orde tinggi GLRLM menggunakan tiga

variable peubah untuk membedakan corak tekstur yaitu nilai aras intensitas *pixel*, sudut dan jarak *pixel* (*Run Length*).

Gambar 2.16 merupakan contoh gambaran pembentukan matriks GLRLM dengan ukuran citra 4x4 dan aras keabuan 0 sampai dengan 3 (4 bit) pada sudut 0° dan 45°. Matriks GLRLM dinyatakan dengan notasi  $p(i, j)$  dari matriks ukuran  $M \times N$  yang tersusun berdasar bari  $i$  aras intensitas keabuan (*Gray Level*) dan kolom  $j$  panjang (*Run Length*) secara berturut-turut pada arah horizontal 0°, diagonal 45°/135° atau vertical 90°. Pada gambar 2.18 nilai intensitas 1 dengan panjang *pixel* 3 pada arah horizontal 0° adalah 1 ( $p(1, 3) = 1$ ).

Picture:		0	1	2	3					
		0	2	3	3					
		2	1	1	1					
		3	0	3	0					
Run length:										
G r a y l e v e l	0°	1	2	3	4	45°	1	2	3	4
	0	4	0	0	0	0	4	0	0	0
	1	1	0	1	0	1	4	0	0	0
	2	3	0	0	0	2	0	0	1	0
	3	3	1	0	0	3	3	1	0	0

Gambar 2.16. Pembentukan matriks GLRLM (Nugroho, 2015)

Konsep tekstur yang diusulkan Galloway disederhanakan oleh Tang menjadi 4 definisi matriks baru sebagai berikut (Nugroho, 2015) :

### 1. Gray Level Run Length Pixel Number Matrix

$$P_p(i, j) = p(i, j) \cdot j \quad (2.45)$$

## 2. Gray Level Run Number Vector

$$p_g(i) = \sum_{j=1}^N p(i, j) \quad (2.46)$$

## 3. Run Length Run Number Vector

$$p_r(j) = \sum_{i=1}^M p(i, j) \quad (2.47)$$

## 4. Gray Level Run Length One Vector

$$p_o(i) = p(i, 1) \quad (2.48)$$

Sebanyak 11 ciri tekstur dapat digali dari elemen matriks GLRLM tersebut dengan rincian sebagai berikut (Nugroho, 2015)

### 1. Short Run Emphasis (SRE)

$$SRE = \frac{1}{n_r} \sum_{j=1}^N \frac{p_r(j)}{j^2} \quad (2.49)$$

dengan  $n_r$  merupakan jumlah total seluruh nilai elemen-elemen matriks GLRLM atau dapat dituliskan sebagai :

$$n_r = \sum_{i=1}^M \sum_{j=1}^N p(i, j) = \sum_{i=1}^M p_g(i) = \sum_{j=1}^N p_r(j) \quad (2.50)$$

2. *Long Run Emphasis (LRE)*

$$LRE = \frac{1}{n_r} \sum_{j=1}^N p_r(j)j^2 \quad (2.51)$$

3. *Gray Level Nonuniformity (GLN)*

$$GLN = \frac{1}{n_r} \sum_{i=1}^M p_g(j)^2 \quad (2.52)$$

4. *Run Length Nonuniformity (RLN)*

$$RLN = \frac{1}{n_r} \sum_{j=1}^N p_r(j)^2 \quad (2.53)$$

5. *Run Percentage (RP)*

$$RP = \frac{n_r}{n_p} \quad (2.54)$$

dengan  $n_p$  merupakan jumlah *pixel* pada matriks GLRLM yang dituliskan sebagai berikut :

$$n_p = M \times N \quad (2.55)$$

6. *Low Gray-level Emphasis (LGRE)*

$$LGRE = \frac{1}{n_r} \sum_{i=1}^M \frac{p_g(i)}{i^2} \quad (2.56)$$

7. *High Gray-level Run Emphasis (HGRE)*

$$HGRE = \frac{1}{n_r} \sum_{i=1}^M p_g(i)(i)^2 \quad (2.57)$$

8. *Short Run Low Gray-level Emphasis (SRLGE)*

$$SRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j)}{i^2 j^2} \quad (2.58)$$

9. *Short Run High Gray-level Emphasis (SRHGE)*

$$SRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j) \cdot i^2}{j^2} \quad (2.59)$$

10. *Low Run Low Gray-level Emphasis (LRLGE)*

$$LRLGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{p(i, j) \cdot j^2}{i^2} \quad (2.60)$$

11. *Low Run High Gray-level Emphasis (LRHGE)*

$$LRHGE = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N p(i, j) i^2 j^2 \quad (2.61)$$

Ciri statistik orde satu, orde dua dan orde tinggi yang dipadukan dalam penelitian diharapkan dapat lebih baik dalam mengenali pola tekstur kanker yang terdapat dalam citra mamogram. Perbedaan frekuensi kemunculan nilai aras intensitas pada citra dianalisis dengan ciri statistik orde satu histogram. Tingkat keseragaman antar *pixel* dari berbagai sudut akan dikenali dengan ciri statistik

orde dua GLCM. Pola tekstur dari berbagai panjang *pixel* di seluruh resolusi citra input akan dikenali oleh ciri statistik orde tinggi GLRLM. Keseluruhan ciri yang digunakan dari ciri statistik orde satu, ciri statistik orde dua dan ciri statistik orde tinggi berjumlah 43 nilai tekstur.

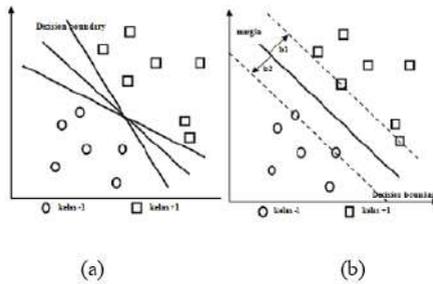
## **7. Seleksi Fitur *Relief***

Seleksi fitur mengacu kepada pemilihan fitur yang optimal dari fitur input yang sesuai dengan kriteria tertentu. Proses seleksi fitur dapat mempercepat proses komputasi dan meningkatkan efisiensi dalam proses klasifikasi (Zhang & Chen, 2018). Salah satu algoritma seleksi fitur adalah *relief*. Metode *relief* berbasis *instance* dimana sampel data diambil secara acak dan sampel terdekatnya dicari pada kelas yang sama dan berlawanan. Hasil estimasi yang diperoleh berdasarkan sampel yang diambil berbeda kemudian dirata-rata untuk mendapat komponen yang relevan dari statistik dari setiap atribut. Komponen yang semakin besar maka kemampuan mengklasifikasi yang sesuai dengan atribut semakin kuat. Langkah pengambilan sampel acak untuk M kali diulangi dan dioperasikan berulang selama N kali, lalu dirata-rata nilai statistik yang relevan dari tiap atribut. Nilai atribut teratas dari komponen yang lebih besar dari statistik yang relevan digunakan untuk proses klasifikasi.

## **8. SVM *Support Vector Machine***

Metode SVM pertama kali dikenalkan pada tahun 1992 oleh Vapnik. SVM (*Support Vector Machine*) merupakan salah satu metode yang banyak digunakan untuk mengklasifikasi citra medis secara

ra otomatis berdasar hasil diagnosa pasien. SVM masuk ke dalam aplikasi *pattern recognition* yang berfungsi memetakan suatu data ke dalam konsep tertentu yang sebelumnya telah didefinisikan dalam bidang komputer sains (Setyorini, 2020). Sistem pembelajaran SVM menggunakan ruang hipotesa berupa fungsi-fungsi lini-er dan dilatih dengan algoritma pembelajaran berdasar teori pembelajaran statistik dalam sebuah ruang ciri (*feature space*). Pada metode SVM selama proses pelatihan hanya sejumlah data terpilih yang akan digunakan dalam proses klasifikasi sehingga ini menjadi salah satu kelebihan SVM (Adi & Isnanto, 2016).



Gambar 2.17. Konsep dasar SVM  
(Adi & Isnanto, 2016)

Gambar 2.17 merupakan ilustrasi konsep SVM. Konsep SVM dalam problem klasifikasi secara sederhana dijelaskan sebagai usaha mencari garis (*hyperlane*) terbaik sebagai pemisah dua buah kelas. Pada gambar 2.17 menunjukkan pemisah dua buah kelas, dimana bentuk lingkaran sebagai simbol untuk kelas -1 dan bentuk persegi merupakan simbol untuk kelas +1. Margin *hyperlane* akan diukur dan dicari titik maksimalnya hingga ditemukan *hyperlane* pemisah terbaik antar dua kelas tersebut. *Hyperlane* terbaik ditunjukkan pada gambar 2.17 (b) karena diantara kedua kelas. Margin sendiri

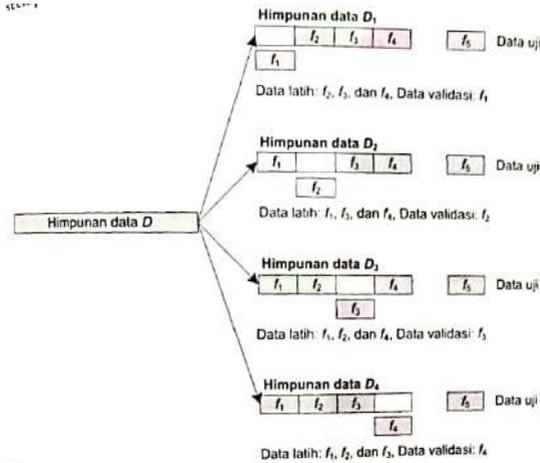
merupakan jarak terdekat data dari masing-masing kelas dengan *hyperlane*. Data terdekat ini disebut sebagai *support vector*. Pada gambar 2.17 (b), *support vector* adalah lingkaran dan persegi yang dilewati garis batas margin (garis putus-putus). Proses pelatihan SVM ini intinya adalah usaha untuk mencari *hyperlane* terbaik (Adi & Isnanto, 2016)

Proses klasifikasi SVM menggunakan bantuan *software machine learning* Weka. Weka merupakan perangkat lunak *open source* untuk *machine learning* berbasis Java. Weka banyak dimanfaatkan untuk pengajaran, penelitian dan aplikasi industry karena memiliki berbagai *tools* bawaan untuk *machine learning* standar. Fasilitas didalam Weka dapat digunakan untuk menganalisa data dan membuat prediksi karena dilengkapi algoritma pembelajaran untuk *data mining*, membantu evaluasi serta membandingkan algoritma-algoritma metode dalam *machine learning*

## 9. *K-Fold cross validation*

Teknik *k-Fold Cross-Validation* (kFCV) merupakan metode validasi dengan membagi himpunan data  $D$  secara acak menjadi  $k$  subhimpunan (biasanya disebut *fold*) yang saling bebas.  $f_1, f_2, \dots, f_k$ , sehingga masing-masing *fold* berisi  $\frac{1}{k}$  bagian data. Selanjutnya  $k$  himpunan data  $D_1, D_2, \dots, D_k$  dibangun dengan masing-masing berisi  $(k - 2)$  *fold* untuk data latih, satu *fold* untuk validasi, dan satu *fold* untuk data uji (Suyanto et al, 2019).

Pada gambar 2.18. merupakan contoh teknik kFCV dengan 5-*fold*. Himpunan data  $D$  dibagi ke dalam 5 *fold* yang saling bebas :  $f_1, f_2, f_3, f_4, f_5$  yang selanjutnya didistribusikan secara sirkular untuk data latih dan data uji. Sebanyak tiga *fold* :  $f_1, f_2, f_3$  untuk data latih,  $f_4$  untuk data validasi dan  $f_5$  untuk data uji di himpun-



Gambar 2.18. Teknik 5-fold cross-validation  
(Suyanto et al,2019)

an data  $D_1$ . Kemudian himpunan data  $D_2$  berisi tiga *fold* :  $f_1, f_2, f_4$  untuk data latih  $f_3$  untuk validasi dan  $f_5$  untuk data uji. Selanjutnya himpunan data  $D_3$  berisi tiga *fold* :  $f_1, f_3, f_4$  untuk data latih  $f_2$  untuk validasi dan  $f_5$  untuk data uji. Terakhir himpunan data  $D_4$  berisi tiga *fold* :  $f_2, f_3, f_4$  untuk data latih  $f_1$  untuk validasi dan  $f_5$  untuk data uji. Pada teknik ini, satu *fold*  $f_5$  sengaja dijadikan untuk data uji. Artinya, data uji hanya satu *fold* tunggal yang tetap. Secara keseluruhan akan menghasilkan empat model DI, seperti diilustrasikan pada gambar diatas. Terakhir, dipilih satu model paling optimum dengan performansi tertinggi dan estimasi model dengan ukuran statistik : performansi minimum, performansi maksimum dan performansi rata-rata (dari keempat model tersebut) (Suyanto et al,2019).

## 10. Confusion Matrix

Pada konsep *data mining* perhitungan akurasi dilakukan dengan metode *confusion matrix*. Terdapat empat istilah yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN) untuk merepresentasikan proses klasifikasi. TN merupakan data negative yang terdeteksi benar, FP merupakan data negative yang terdeteksi sebagai data positif, TP merupakan data positif yang terdeteksi benar dan FN merupakan data positif yang terdeteksi sebagai data negative (Setyorini, 2020)

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Gambar 2.19. *Confusion matrix* (Setyorini, 2020)

Berdasarkan nilai *confusion matrix* seperti pada gambar 2.19 maka indeks pengukuran pada proses klasifikasi dapat dihitung. Indeks pengukuran tersebut antara lain:

### 1. Akurasi

$$Akurasi(Accuration) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.62)$$

### 2. Presisi Menggambarkan jumlah data kategori positif yang diklasifikasi secara benar dibagi total data yang diklasifikasi

positif

$$Presisi(precision) = \frac{TP}{TP + FP} \times 100\% \quad (2.63)$$

3. Menunjukkan persentase data kategori positif yang terklasifikasi dengan benar oleh sistem.

$$Recall(sensitivitas) = \frac{TP}{TP + FN} \times 100\% \quad (2.64)$$

4. *Specificity* Merupakan nilai kebenaran dari prediksi negative dibanding dengan seluruh data negative.

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (2.65)$$

## B. Kajian Pustaka

Beberapa penelitian mengenai klasifikasi citra medis berbasis *Computer Diagnosis Aided* (CAD) menggunakan ekstraksi fitur tekstur telah banyak dilakukan. Kadek dan I Made melakukan klasifikasi citra *mammogram* dengan metode segmentasi K-Means, metode ekstraksi dengan *Gray Level Co-Occurrence Matrix* (GLCM) dengan mengambil 4 nilai ciri dan *Support Vector Machine* (SVM) sebagai proses klasifikasi. Klasifikasi citra *mammogram* dibedakan menjadi kelas normal dan tidak normal dan didapatkan akurasi sebesar 80% (Setiawan, 2018).

Penelitian untuk klasifikasi kanker payudara yang dilakukan oleh Laksono menunjukkan hasil akurasi sebesar 87,57% pada sistem klasifikasi SVM. Citra yang digunakan adalah citra *mammogram* dengan proses segmentasi menggunakan *Region Growing*. Selanjutnya citra di ekstraksi fitur menggunakan GLCM (*gray level co-occurrence*

*matrix*) dengan mengambil sebanyak 13 ciri. Setelah citra di ekstraksi kemudian di klasifikasi dengan menggunakan SVM *Support Vector Machine* (Laksono, 2017)

Penelitian yang dilakukan Lussiana menggunakan citra *mammogram* untuk diklasifikasi sebagai kelas normal dan tidak normal. Metode segmentasi dengan menggunakan K-Means dan selanjutnya ditransformasikan menggunakan Wavelet Daubechies 6 sampai dengan dekomposisi level 2. Citra kemudian di ekstraksi untuk menghasilkan nilai-nilai yang menjadi parameter untuk klasifikasi. Proses ekstraksi ciri menggunakan GLCM dengan 3 ciri. Klasifikasi yang digunakan dengan metode *Support Vector Machine* (SVM) dan didapat akurasi 91% untuk kategori normal dan 87.67% untuk abnormal (Lussiana dkk, 2013).

Penelitian oleh Rosadi untuk mengklasifikasi masa pada citra *mammogram*. Metode ekstraksi ciri dengan menggunakan GLCM kemudian di seleksi fitur dengan F-Score untuk selanjutnya diklasifikasi dengan metode SVM sebagai kelas kanker jinak dan kanker ganas. Hasilnya menunjukkan bahwa SVM dengan fungsi kernel RBF mampu menghasilkan akurasi sebesar 97% untuk proses klasifikasi (Rosadi, 2016).

Penelitian berikutnya oleh Nugroho yaitu untuk mengklasifikasikan nodul tiroid pada citra USG berdasarkan tekstur menjadi 2 kelas, kistik dan solid. Pembuatan sistem menggunakan software MATLAB versi 2011a. Proses peningkatan kualitas citra menggunakan tapis median, proses ekstraksi menggunakan 9 ciri dari Histogram, 21 ciri dari *Grey Level Co-occurrence Matrices* (GLCM) dan 11 ciri dari *Gray Level Run Length Matrices* (GLRLM), seleksi ciri menggunakan *information gain* dengan hasil seleksi ciri sebanyak 22 ciri. Proses klasifikasi menggunakan MLP menunjukkan

bahwa metode yang telah dilakukan dapat mencapai akurasi klasifikasi 86,1%, sensitivitas 89,5%, spesifisitas 82,4%. Penelitian ini menyatakan bahwa dengan melakukan penambahan ciri tekstur untuk meningkatkan keakuratan sistem *recognition* dengan tetap memprioritaskan ciri-ciri tekstur yang unik dan dominan (Nugroho, 2015).

Penelitian oleh Eli untuk mengklasifikasi nodul payudara menggunakan ekstraksi ciri tekstur dengan 9 ciri histogram dan 21 ciri *Grey Level Co-occurrence Matrices* (GLCM). Hasil Ekstraksi menggunakan Scilab menunjukkan bahwa dengan menggunakan 30 ciri tekstur, *Multi Layer Perceptron* (MLP) dapat mengklasifikasikan nodul bermassa kistik dan solid dengan akurasi sebesar 88,89% dan 80,56% menggunakan 10 ciri ciri tekstur (Ermawati, 2020).

Penelitian-penelitian sebelumnya telah berhasil membangun suatu sistem pengolahan citra medis berbasis sistem CAD. Tiga diantaranya menggunakan *Support Vector Machine* (SVM) sebagai proses klasifikasi dan mendapatkan hasil akurasi yang baik bahkan mencapai diatas 90%. Penelitian-penelitian dengan SVM sebagai metode klasifikasi hanya menggunakan satu dari tiga analisis fitur tekstur statistik yaitu GLCM. Penelitian juga dilakukan menggunakan *software* Matlab. Sementara pada penelitian ini, sistem CAD yang akan dibangun terfokus pada proses klasifikasi kanker jinak dan kanker ganas berdasarkan tiga jenis fitur tekstur statistik yaitu histogram, GLCM dan GLRLM menggunakan *software* Scilab.

## **BAB III**

### **METODE PENELITIAN**

#### **A. Waktu dan Tempat Penelitian**

Penelitian dalam proses pengolahan data, analisis data dan pembuatan laporan akan dilaksanakan pada 1 Agustus 2021 – 1 Mei 2022 di Laboratorium Komputasi Jurusan Fisika Fakultas Sains dan Teknologi UIN Walisongo Semarang.

#### **B. Alat Penelitian**

Alat yang digunakan pada penelitian ini adalah :

1. Laptop dengan spesifikasi processor Intel® Core™ i5-3320M CPU @2,60GHz (4 CPUs), RAM 4GB dan sistem operasi Windows 10 Pro 64-bit (10.0, Build 19042).
2. Software Scilab versi 6.0.2 untuk melakukan preprocessing dan ekstraksi ciri.
3. Machine learning Weka versi 3.8.4 untuk melakukan klasifikasi.
4. Texmaker versi 5.0.4 untuk membuat laporan penelitian.

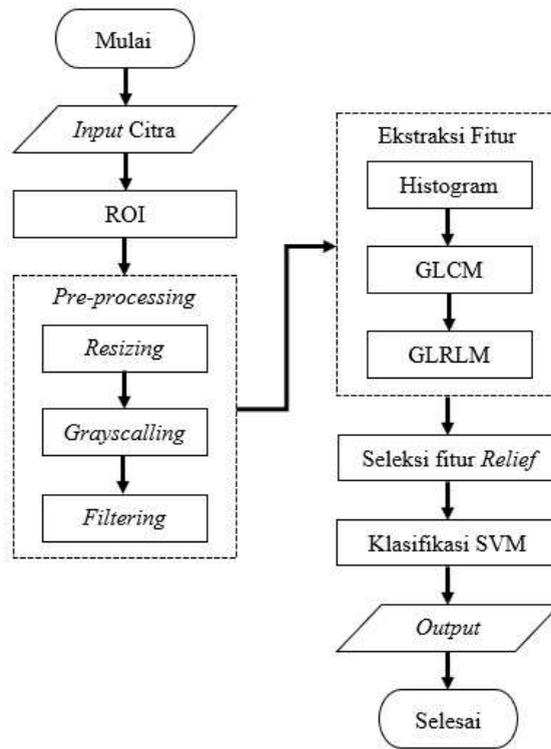
#### **C. Bahan Penelitian**

Bahan dalam penelitian ini adalah data berupa citra digital mamografi dari *The Mammographic Image Analysis Society* (MIAS) yang diakses dan diunduh melalui laman <http://www.mammoimage.org/>

databases/. Citra mammografi berasal dari organisasi kelompok penelitian Inggris. Citra digital dengan resolusi 1024 x 1024 pixel dan format gambar PGM diambil dari UK *National Breast Screening Programme* pada 2015. Jumlah keseluruhan data sebanyak 322 citra digital dengan informasi area yang dicurigai sebagai tumor jinak (*benigna*) atau tumor ganas (*malignant*) dalam koordinat x dan y oleh ahli radiologi. Namun data yang diambil untuk penelitian ini yaitu tumor jinak sebanyak 42 citra dan tumor ganas sebanyak 25 citra. Jumlah tersebut dipilih karena citra dengan derajat kehalusan yang paling baik berdasarkan pada informasi klinis yang terdapat dalam database. Citra tumor jinak berjumlah lebih banyak agar pada saat proses klasifikasi *machine learning* mampu bekerja mengenali data dengan baik (Laksono, 2017)

#### **D. Pengolahan Data**

Prosedur penelitian ini akan dilakukan dengan beberapa tahapan pengolahan data seperti ditunjukkan pada gambar 3.1 berikut



Gambar 3.1. Tahapan pengolahan data

1. Input citra Citra input merupakan citra mammografi yang terdiri dari dua kelas yaitu citra tumor jinak dan citra kanker. Citra masih terdapat derau dan marker dari database dengan arah pengambilan hanya dari samping payudara atau *MLO (Medio-Lateral Oblique)*. Seluruh citra merupakan citra RGB berukuran 1024 x 1024 *pixel* dengan format PNG.
2. RoI  
Langkah awal dalam pengolahan citra yaitu RoI (*Region of In-*

*terest*) yang bertujuan menentukan lokasi kanker pada citra. Lokasi kanker telah ditandai pada *database* oleh ahli radiologi kemudian di *cropping* untuk proses selanjutnya.

### 3. *Pre-processing*

Tahapan *Pre-processing* dilakukan untuk meningkatkan kualitas citra. Proses pertama yaitu proses *resizing* atau mengubah ukuran citra menjadi 256 x 256 *pixel*. Citra kemudian di-konversi menjadi citra berskala keabuan (*grayscale*). Tahap selanjutnya yaitu dilakukan proses *filtering* pada citra untuk menghilangkan derau dan *marker* pada citra dengan menggunakan *median filter*.

### 4. Ekstraksi fitur

Tahapan ke empat yaitu ekstraksi fitur (penggalan ciri) pada citra. Fitur yang dicari dalam tahapan ekstraksi ini adalah fitur tekstur (statistik) dengan tiga jenis fitur statistik yaitu histogram, GLCM dan GLRLM. Parameter fitur tekstur histogram yang dihitung berjumlah 9 antara lain *mean*, *variance*, entropi dan standar deviasi. Parameter untuk fitur tekstur GLCM berjumlah 24 antara lain yaitu *contrast*, *correlation*, *uniformity* dan *homogeneity*. Untuk fitur tekstur GLRLM parameter yang dihitung berjumlah 11 antara lain adalah *Short Run Emphasis* (SRE), *Long Run Emphasis* (LRE), *Gray Level Nonuniformity* (GLN), *Run Length Nonuniformity* (RLN) dan *Run Percentage*(RP). Hasil dari ekstraksi fitur berupa parameter tersebut akan memudahkan dalam proses klasifikasi citra

### 5. Seleksi fitur

Seleksi fitur dilakukan guna mendapatkan atribut-atribut pada citra yang mempengaruhi proses klasifikasi. Seleksi fitur yang digunakan adalah *Relief*. Teknik pemeringkatan atribut dengan *relief* menggunakan sample acak dari sebuah data dan tetangga terdekat dari kelas yang sama atau berlawanan dicari. Kemudian nilainya akan dibandingkan dengan sampel *instance* untuk diperbaharui skor relevansi dari tiap atribut

## 6. Klasifikasi

Tahapan yang terakhir adalah proses klasifikasi data dengan menggunakan *Support Vector Machine* (SVM). Proses pelatihan SVM, merupakan upaya untuk mencari hyperlane pemisah terbaik antar dua kelas. SVM menggunakan ruang hipotesa berupa fungsi linier dan dilatih dengan algoritma pembelajaran. *K-Fold cross validation* dipilih sebagai algoritma pembelajaran yang membagi data menjadi data latih (*training-set*) dan data uji (*testingset*). Proses klasifikasi menggunakan bantuan *machine learning weka* dan data dikelompokkan menjadi dua kategori kelas yaitu kanker jinak dan kanker ganas.

## E. Analisis Data

Analisis data dilakukan pada proses klasifikasi dengan tujuan untuk mengetahui tingkat akurasi proses klasifikasi berdasarkan ciri tekstur yang digunakan. Analisis hasil klasifikasi dilakukan dengan perhitungan indeks pengukuran seperti : akurasi, presisi, sensitivitas dan *specificity* berdasarkan hasil *confusion matrix* dari *machine learning Weka*.

Nilai akurasi menggambarkan keakuratan metode dalam melakukan klasifikasi pada seluruh data dengan benar. Nilai sensitivitas menggambarkan keberhasilan metode klasifikasi dalam menemukan informasi data kembali. Nilai spesififikasi menunjukkan kebenaran metode dalam memprediksi data negatif yang dibandingkan dengan seluruh data negatif. Nilai PPV menunjukkan ketepatan metode untuk suatu kategori sebagai kategori itu sendiri. Nilai NPV menunjukkan ketepatan metode untuk mengklasifikasi non-kategori sebagai non-kategori itu sendiri.

Hasil analisis data pada penelitian ini digunakan untuk menilai performa metode yang digunakan dalam proses klasifikasi. Performa metode klasifikasi dengan nilai akurasi tertentu dikategorikan menjadi beberapa kelas kualitas berdasarkan ROC (*Receiver Operating Characteristic*). Rentang akurasi 90% - 100% merupakan kategori *excellent*. Akurasi 80% - 90% kategori *good*. Akurasi 70% - 80% kategori *fair*. Akurasi 60% - 70% kategori *poor* dan akurasi 50% - 60% merupakan kategori *failure* (Hariati, 2018)

## BAB IV

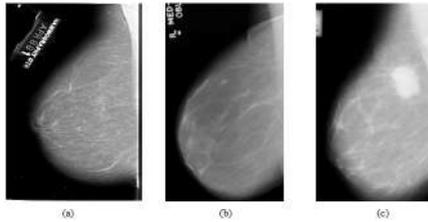
### HASIL DAN PEMBAHASAN

Penelitian ini dibagi menjadi beberapa tahapan yang akan dipaparkan hasil serta pembahasannya pada bab ini.

#### A. *Pre-Processing*

Tahap pertama pada penelitian ini yaitu *pre-processing* data, dengan tujuan untuk memperbaiki kualitas citra sebelum dilakukan ekstraksi fitur. Data citra yang digunakan didapat dari *Mammographic Image Analysis Society* (MIAS) tahun 2015. MIAS merupakan organisasi dari kelompok penelitian Inggris. Basis data berisi 322 citra digital dengan ukuran yang dibuat sama besar yaitu 1024 x 1024 pixel. Karena citra yang digunakan berukuran cukup besar, maka perlu dilakukan pemotongan atau *crop* sehingga bagian yang digunakan adalah difokuskan pada area tumor payudara yang disebut RoI (*Region of Interest*). Proses *cropping* dilakukan dengan menjadikan keterangan titik koordinat X dan Y lokasi tumor serta ukuran tumor pada informasi klinis database sebagai acuan. Citra hasil *cropping* yang disebut sebagai citra RoI memiliki ukuran lebih kecil dari citra asli yaitu 256 x 256 pixel. Citra RoI kemudian disimpan dalam format JPG. Pada penelitian ini, data citra yang digunakan sebanyak 67 citra, dengan citra tumor jinak sebanyak 42 citra dan tumor ganas sebanyak 25 citra. Jumlah tersebut dipilih karena citra dengan derajat kehalusan yang paling baik berdasarkan pada informasi klinis yang terdapat dalam database. Citra tumor jinak berjumlah lebih banyak agar saat proses klasifikasi *machine learning* mampu bekerja mengenali data dengan baik. Gambar 4.1.

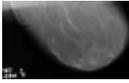
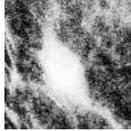
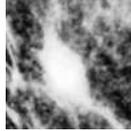
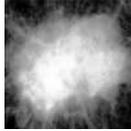
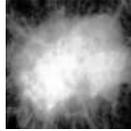
memperlihatkan citra payudara normal, tumor jinak (*benigna*) dan tumor ganas (*malignant*) pada payudara.



Gambar 4.1. Citra(a)normal (b)*benigna* (c)*malignant*  
(Berment et al., 2014)

Setelah citra asli di cropping, citra kemudian dilakukan ekualisasi citra menggunakan ekualisasi histogram. Ekualisasi histogram bertujuan untuk mengatur tingkat kecerahan pada citra sesuai dengan nilai kecerahan histogram tetangga terdekatnya. Sehingga citra yang didapat merupakan citra dengan nilai derajat keabuan yang rata. Proses selanjutnya, citra dilakukan *filtering* dengan menggunakan median adaptif filter dengan tujuan mengurangi derau (*noise*) yang terbawa dalam proses pengambilan citra. Proses ekualisasi citra dan *filtering* pada penelitian ini diselesaikan dengan menggunakan *software* scilab. Citra hasil RoI serta *pre-processing* dapat dilihat pada Tabel 4.1.

Tabel 4.1. Citra Hasil Ekualisasi Histogram dan Median Adaptif Filter

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori
				<i>benigna</i>
				<i>malignant</i>

Pada tahap *pre-processing* seperti pada tabel 4.1 terlihat peningkatan kualitas citra secara visual. Batas antara area tumor dengan jaringan sekitarnya terlihat lebih jelas. Area tumor payudara pada citra memiliki intensitas lebih terang dibandingkan area jaringan payudara yang memiliki intensitas lebih gelap. Citra hasil *pre-processing* kemudian disimpan dalam format JPG untuk selanjutnya dilakukan proses ekstraksi citra.

## B. Ekstraksi Fitur

Tahap ekstraksi fitur merupakan proses penggalian ciri pada semua citra hasil *pre-processing*. Fitur tekstur yang digunakan terdiri dari 9 parameter histogram, 24 parameter GLCM dan 11 parameter GLRLM. Tahap ekstraksi fitur diselesaikan dengan algoritma komputasi *scilab* dan menghasilkan data kuantitatif berupa nilai-nilai dari hasil perhitungan setiap parameter. Data yang didapat berupa file *notepad* (.txt) yang kemudian di ubah menjadi file excel

(.xls) untuk mempermudah pembacaan hasil.

Tabel 4.2, tabel 4,3 serta tabel 4,4 merupakan data hasil ekstraksi fitur tekstur. Data yang ditampilkan merupakan nilai rata-rata dari masing-masing tekstur. Untuk data GLCM dan GLRLM data merupakan rata-rata nilai dari ke empat sudutnya yaitu sudut  $0^\circ, 45^\circ, 90^\circ, 135^\circ$ . Data dengan nilai rata-rata tidak dipergunakan pada proses klasifikasi. Data hasil secara keseluruhan terdapat pada lampiran 2. Data keseluruhan yang kemudian digunakan pada tahap klasifikasi. Seluruh data sebelumnya sudah dikonversi menjadi data dengan format CSV untuk mempermudah pembacaan data pada proses selanjutnya.

Tabel 4.2. Rata-rata Hasil Ekstraksi Fitur Tekstur Histogram

No	Fitur	Tumor Jinak	Tumor Ganas
1	Mean	$1,000 \pm 0,000$	$1,000 \pm 0,000$
2	Modus	$1851,548 \pm 1633,502$	$2137,800 \pm 3442,067$
3	Energi	$2,000 \pm 0,000$	$2,000 \pm 0,000$
4	Entropi	$5,430 \pm 0,149$	$5,477 \pm 0,049$
5	Skewness	$65,266 \pm 1,853$	$64,926 \pm 0,621$
6	Kurtosis	$13294,773 \pm 462,087$	$13240,778 \pm 163,071$
7	Varians	$22200,640 \pm 469,070$	$22059,356 \pm 182,136$
8	Dev Std	$145,800 \pm 20,521$	$148,523 \pm 0,612$
9	Smoothness	$0,254 \pm 0,004$	$0,253 \pm 0,002$

Tabel 4.3. Rata-rata Hasil Ekstraksi Fitur Tekstur GLCM

No	Fitur	Tumor Jinak	Tumor Ganas
1	Contrast	131,001 ± 25,677	18,741 ± 10,709
2	ASM	0,001 ± 0,001	0,001 ± 0,000
3	Acorr	22751,079 ± 477,73	22637,942 ± 173,099
4	Dissi	3,391 ± 1,255	2,728 ± 0,765
5	Inertia	31,001 ± 25,677	18,741 ± 10,709
6	Joint Average	132,405 ± 2,007	131,668 ± 1,018
7	Joint Entropi	8,042 ± 0,280	7,950 ± 0,223
8	Joint Variance	22766,575 ± 481,857	22647,308 ± 175,978
9	Correlation	0,000 ± 0,000	0,000 ± 0,000
10	Variance	5215,796 ± 271,6124	5300,404 ± 152,986
11	Cshade	-96311,816 ± 111813,751	-74781,162 ± 84806,081
12	Cprom	8,631 ± 0,012	8,317 ± 0,013
13	Mprob	0,005 ± 0,008	0,004 ± 0,003
14	SAV	264,116 ± 4,085	262,682 ± 2,158
15	SEntro	6,145 ± 0,129	6,175 ± 0,030
16	Svar	87472,242 ± 1969,929	87029,37 ± 739,406
17	DEntro	2,272 ± 0,297	2,107 ± 0,241
18	DVar	30,974 ± 25,667	18,720 ± 10,704
19	INV	0,423 ± 0,058	0,455 ± 0,055
20	INN	108,378 ± 14,867	116,576 ± 14,307
21	IDM	0,353 ± 0,642	0,388 ± 0,064
22	IDMN	27744,989 ± 3806,179	29843,555 ± 3662,669
23	Imcorr 1	2-0,518 ± 0,054	-0,547 ± 0,048
24	Imcorr 2	0,997 ± 0,001	0,998 ± 0,001

Tabel 4.4. Rata-rata Hasil Ekstraksi Fitur Tekstur GLRLM

No	Fitur	Tumor Jinak	Tumor Ganas
1	SRE	0,8924 ± 0,029	0.884 ± 0.028
2	LRE	1,851 ± 0,432	1,930 ± 0,426
3	GLN	266,087 ± 69,258	238,680 ± 27,205
4	RLN	41853,367 ± 5363,491	40403,824 ± 5212,294
5	RP	0,837 ± 0,048	0,825 ± 0,048
6	LGRE	0,001 ± 0,000	0,001 ± 0,000
7	HGRE	22112,904 ± 969,713	22160,640 ± 891,234
8	SRLGE	0,001 ± 0,000	0,0007 ± 0.0002
9	SRHGE	19558,851 ± 1026,409	19426,373 ± 1058,105
10	LRLGE	0,014 ± 0,015	0,011 ± 0,011
11	LRHGE	43230,360 ± 8206,769	45457,262 ± 9012,637

### C. Seleksi Fitur

Tahap seleksi fitur merupakan tahap yang bertujuan untuk mengetahui fitur tekstur yang berpengaruh dalam proses klasifikasi. Seleksi fitur menggunakan metode seleksi fitur *Relief* dengan bantuan *machine learning* Weka. Metode ini bekerja dengan mengambil nilai atribut teratas untuk kemudian digunakan pada proses klasifikasi. Tabel 4.5 merupakan rangkuman hasil seleksi fitur dengan urutan nilai atribut teratas hingga terbawah. Nilai atribut yang akan digunakan nantinya merupakan nilai dengan tidak sama dengan nol dan juga bernilai positif. Sementara nilai atribut dengan nilai sama dengan nol dan nilai negatif tidak akan digunakan karena tidak berpengaruh pada proses klasifikasi. Berdasarkan hasil seleksi fitur terdapat 128 parameter dari total 149 parameter yang berpengaruh dalam klasifikasi tumor payudara.

Tabel 4.5. Hasil seleksi fitur tekstur

No	Parameter Fitur	Ranked
1	IDM135	0,0185
2	INN90	0.0183
3	IDMN135	0.0183
4	INV90	0.0183
5	INNO	0.0183
6	RLNO	0.0182
7	SREO	0.0176
...	...	...
129	ENERGI	0
130	MEAN	0
...	...	...
148	LRLGEO	-0.0073
149	DEV STD	-0.0080

#### D. Klasifikasi

Tahap klasifikasi merupakan tahap akhir yang dilakukan dalam pengolahan citra. Klasifikasi menggunakan jaringan *Support Vector Machine* (SVM) dengan bantuan *machine learning* Weka seperti pada gambar 4.2. Data masukan yang digunakan merupakan data hasil pada tahap ekstraksi fitur dengan format CSV sehingga memudahkan pembacaan data pada *machine learning* Weka. Data masukan yang digunakan pada proses klasifikasi merupakan seluruh data gabungan hasil ekstraksi fitur tekstur yaitu histogram, GLCM dan GLRLM dengan jumlah fitur total sebanyak 149 fitur. Data masukan digunakan sebagai data latih pada *trainingset* juga sebagai data uji pada *testingset* dengan metode *10-fold cross validation*. Metode ini bekerja dengan membagi data masukan klasifikasi menjadi 10-fold sehingga diperoleh 10 dataset untuk mengevaluasi kinerja

proses klasifikasi.



Gambar 4.2. Tampilan awal pada *machine learning* Weka

Klasifikasi dilakukan dengan 3 kali proses klasifikasi yaitu *trainingset*, *testingset* dan data hasil seleksi fitur. Proses klasifikasi *trainingset* berguna untuk mengetahui apakah jaringan SVM yang digunakan pada proses klasifikasi sudah mampu mengenali setiap data dengan baik. Semakin tinggi nilai akurasi pada proses *trainingset* menandakan bahwa SVM memiliki kinerja yang baik. Sementara proses klasifikasi *testingset* merupakan hasil klasifikasi yang sesungguhnya untuk pengujian SVM. *testingset* merupakan hasil untuk mengetahui kebenaran dari proses pelatihan SVM.

Proses klasifikasi dinilai berhasil berdasarkan indeks pengukuran *confusion matrix* dimana terdiri dari *True Positive* (TP), *False Positive* (FP), *False Negative* (FN) dan *True Negative* (TN). TP merupakan citra dengan kategori *benigna* yang terklasifikasi *benigna* pada sistem. FP merupakan citra dengan kategori *malignant* yang terklasifikasi *benigna*. TN merupakan citra dengan kategori *malignant* yang terklasifikasi *malignant* pada sistem dan FN merupakan citra dengan kategori *benigna* yang terklasifikasi *malignant* pada sistem.

```

=== Confusion Matrix ===
  a  b  <-- classified as
42  0  |  a = B
 0 25  |  b = M

```

Gambar 4.3. *Confusion Matrix* Hasil Klasifikasi *Trainingset*

```

=== Confusion Matrix ===
  a  b  <-- classified as
42  0  |  a = B
25  0  |  b = M

```

Gambar 4.4. *Confusion Matrix* Hasil Klasifikasi *Testingset*

```

=== Confusion Matrix ===
  a  b  <-- classified as
42  0  |  a = B
25  0  |  b = M

```

Gambar 4.5. *Confusion Matrix* Hasil Klasifikasi Data Hasil Seleksi Ciri

Nilai *Confusion Matrix* selanjutnya digunakan untuk menghitung parameter indeks pengukuran klasifikasi yang kemudian dijadikan penilaian performa hasil metode klasifikasi pada penelitian ini. Perbandingan hasil indeks pengukuran proses klasifikasi ditampilkan pada tabel 4.6.

Tabel 4.6. Hasil seleksi fitur tekstur

Parameter	Hasil <i>trainingset</i>	Hasil <i>testingset</i>	Hasil Seleksi
Jumlah data	67	67	67
Jumlah fitur	149	149	128
TP	42	42	42
FN	0	0	0
FP	0	25	25
TN	25	0	0
<i>Accuracy</i>	100%	62,68%	62,68%
<i>Sensitivity</i>	100%	100%	100%
<i>Specificity</i>	100%	0%	0%
PPV	100%	62,68%	62,68%
NPV	100%	0%	0%

Tabel 4.6 menunjukkan bahwa nilai akurasi, sensitivitas, spesifisitas, PPV serta NPV pada proses *trainingset* bernilai masing-masing 100%. Nilai 100% diperoleh karena dari total 67 data *trainingset* dapat dihasilkan 67 data yang terklasifikasi dengan benar yaitu 42 TP dan 25 TN tanpa adanya data yang terklasifikasi salah pada FP dan FN. Ini menunjukkan bahwa SVM sukses mengenali data latih citra *benigna* dan *malignant* berdasarkan hasil ekstraksi fitur tekstur yang belum terseleksi.

Hasil berbeda diperoleh pada hasil *testingset* dimana nilai akurasi, sensitivitas, spesifisitas, PPV serta NPV berturut-turut adalah 62,68%, 100%, 0%, 62,68% dan 0%. Nilai sensitivitas sebesar 100% artinya kemampuan klasifikasi SVM untuk menilai citra *benigna* pada data dengan kategori *benigna* adalah sebesar 100%. Nilai spesifisitas sebesar 0% artinya kemampuan SVM untuk menilai citra *malignant* pada data dengan kategori *malignant* sebesar 0%. Sementara nilai akurasi sebesar 62,68% merupakan kemampuan SVM dalam mengklasifikasi secara benar seluruh data yang

diuji sebesar 62,68%. Nilai-nilai tersebut diperoleh karena dari total 67 data *testingset* dihasilkan 42 data yang terklasifikasi dengan benar yaitu 42 TP dan 0 FN.

Nilai spesifisitas dan NPV yang rendah diakibatkan karena pada proses klasifikasi SVM didapatkan nilai FP (*false positive*) yang tinggi yaitu sebanyak 25 FP. Artinya, data yang seharusnya terklasifikasi *malignant* namun terdeteksi *benigna* oleh sistem SVM. Sementara untuk mendapatkan nilai spesifisitas dan NPV yang tinggi, nilai *false positive* yang didapatkan harus rendah. Untuk meningkatkan efisiensi dalam proses klasifikasi dilakukan seleksi fitur *relief* dengan dilakukan pemilihan atribut yang berpengaruh dalam proses klasifikasi. Dari total 149 fitur data, didapatkan sebanyak 128 fitur yang berpengaruh dalam proses klasifikasi. Namun, dalam penelitian ini hasil klasifikasi sebelum dan setelah seleksi fitur adalah sama yaitu diperoleh nilai akurasi, sensitivitas, spesifisitas, PPV serta NPV berturut-turut adalah 62,68%, 100%, 0%, 62,68% dan 0%. Hal ini menunjukkan bahwa seleksi fitur *relief* tidak berpengaruh terhadap nilai hasil pada proses klasifikasi SVM.

Hasil klasifikasi yang diperoleh pada penelitian ini menunjukkan nilai akurasi yang rendah apabila dibandingkan dengan penelitian sebelumnya. Penelitian Kadek dan I Made (2018), Laksono (2017) dan Lussiana dkk (2013) mampu menghasilkan nilai akurasi lebih dari 80% dengan menggunakan klasifikasi SVM. Perbedaan hasil klasifikasi disebabkan oleh beberapa faktor seperti metode yang digunakan dalam pengolahan citra digital berbasis CAD. Hasil tahap klasifikasi dipengaruhi oleh metode ekstraksi tekstur. Ketiga penelitian sebelumnya hanya menggunakan satu metode ekstraksi tekstur fitur yaitu GLCM.

Penelitian ini berbeda dengan penelitian sebelumnya, dimana

pada penelitian ini menggunakan tiga analisis fitur tekstur yaitu histogram, GLCM dan GLRLM. Sehingga hasil pada proses klasifikasi dipengaruhi oleh ketiga fitur tekstur tersebut. Namun, hasil akurasi yang didapat dibawah 80%. Selain itu, pada penelitian sebelumnya proses klasifikasi menggunakan SVM hanya membagi data menjadi dua kelas citra mammogram normal dan tidak normal. Artinya citra hanya dibedakan apakah citra tersebut citra payudara normal atau citra payudara dengan tumor. Sedangkan pada penelitian ini, kelas citra yang di uji adalah kelas citra payudara dengan tumor jinak (*benigna*) dan tumor ganas (*malignant*) yang secara visual sulit untuk dibedakan. Citra masukkan pada penelitian penelitian sebelumnya tidak melalui RoI (*Region of Interest*) pada proses paling awal *preprocessing* yang memfokuskan citra hanya pada area tumor saja. Pada penelitian sebelumnya citra masukkan merupakan citra asli mammogram payudara yang kemudian di tingkatkan kontras serta dilakukan *filtering*. Proses *preprocessing* ini yang mempengaruhi nilai-nilai perhitungan pada ekstraksi fitur sehingga didapat akurasi yang rendah.

Penelitian dengan menggunakan tiga metode ekstraksi tekstur fitur Histogram, GLCM dan GLRLM pernah dilakukan oleh Nugroho (2015). Penelitian tersebut menggunakan tiga metode tekstur untuk mengklasifikasi nodul tiroid pada citra ultrasonografi. Proses klasifikasi dilakukan dengan menggunakan MLP (*Multi Layer Perceptron*) dan didapat akurasi sebesar 86,1%. Meskipun metode ekstraksi yang digunakan sama namun pada proses klasifikasi pada penelitian ini berbeda, dimana pada penelitian ini menggunakan SVM (*Support Vector Machine*) sehingga hasil yang diperoleh juga berbeda. Selain itu data citra yang digunakan juga berbeda. Penelitian ini menggunakan citra mammografi dan pada pro-

ses klasifikasinya mendapatkan nilai *false positive* (FP) yang cukup tinggi. Sistem SVM belum mampu mengklasifikasi citra dengan kategori *malignant* dengan baik padahal untuk citra dengan kategori *benigna* mampu mengklasifikasi secara benar dengan baik. Hal ini dapat terjadi karena citra dengan kategori *malignant* yang digunakan pada penelitian ini masih merupakan kategori stadium rendah sehingga bentuk citra masih serupa.

Nilai akurasi keseluruhan sebesar 62,68% yang didapat dari perhitungan *confusion matrix* merupakan penilaian kinerja dari *Support Vector Machine* (SVM) dalam mengklasifikasi data. Performa metode klasifikasi dengan nilai akurasi tertentu dikategorikan menjadi beberapa kelas kualitas berdasarkan ROC (*Receiver Operating Characteristic*). Rentang akurasi 90% - 100% merupakan kategori *excellent*. Akurasi 80% - 90% kategori *good*. Akurasi 70% - 80% kategori *fair*. Akurasi 60% - 70% kategori *poor* dan akurasi 50% - 60% merupakan kategori *failure* (Hariati, 2018). Berdasarkan nilai akurasi yang diperoleh sebesar 62,68% artinya kemampuan *Support Vector Machine* (SVM) dalam mengklasifikasi citra mammo-gram berbasis ciri tekstur statistik pada penelitian ini merupakan kategori *poor classification*.

## BAB V

### PENUTUP

#### A. Kesimpulan

1. Proses pengolahan citra mammogram menggunakan metode *Support Vector Machine* (SVM) berdasarkan ciri tekstur statistik diawali dengan RoI (*Region of Interest*) untuk memfokuskan citra pada area tumor payudara. Selanjutnya citra di *pre-processing* untuk meningkatkan kecerahan serta menghilangkan derau dengan ekualisasi histogram dan median adaptif filter. Citra kemudian di ekstraksi fitur tekstur menggunakan tekstur statistik histogram, *Gray Level Co-Occurrence Matrix* (GLCM) dan *Gray Level Run Length Matrices* (GLRLM). Hasil ekstraksi fitur tekstur kemudian di seleksi menggunakan fitur *relief* untuk mencari nilai-nilai fitur yang berpengaruh pada proses klasifikasi. Terakhir, citra diklasifikasi menggunakan metode *Support Vector Machine* (SVM).
2. Proses pengujian klasifikasi tumor payudara pada citra mammogram berdasarkan ciri tekstur statistik histogram, *Grey Level Co-occurrence Matrices* (GLCM) dan *Gray Level Run Length Matrices* (GLRLM) menggunakan metode *Support Vector Machine* (SVM) diperoleh hasil akurasi sebesar 62,68%. Berdasarkan ROC (*Receiver Operating Characteristic*) untuk penilaian performa metode klasifikasi maka termasuk kategori *poor classification*

## B. Saran

Berikut ini merupakan beberapa saran agar dapat melengkapi penelitian ini untuk penelitian selanjutnya :

1. Hasil ekstraksi ciri tekstur sangat dipengaruhi oleh kualitas citra masukan sehingga diperlukan tambahan algoritma *software* scilab pada tahap *pre-processing* agar citra masukan memiliki kualitas yang lebih baik.
2. Tahap klasifikasi tumor payudara dapat dilakukan dengan metode lain sehingga diharapkan dapat meningkatkan hasil akurasi.
3. Masih ada parameter lain untuk melakukan identifikasi tumor payudara pada citra mammogram yang dapat dipelajari pada tulisan ilmiah kedokteran.

## DAFTAR PUSTAKA

- Adi, K., & Isnanto, R. R. 2016. *Support Vector Machine Untuk Klasifikasi Citra Jenis Daging Berdasarkan Tekstur Menggunakan Ekstraksi Ciri Gray Level Co-Occurrence Matrices ( GLCM )*. *Jurnal Sistem Informasi Bisnis* 1-10
- Baudin, M. 2010. *Introducing to Scilab 1-14*
- Berment, H., Becette, V., Mohallem, M., Ferreira, F., & Chérel, P. 2014. *Masses in mammography: What are the underlying anatomopathological lesions? Diagnostic and Interventional Imaging* 95(2), 124–133
- Chityala, R., & Pudipeddi, S. 2014. *Image Processing and Acquisition using Python*. New York : CRC Press.
- Dewi, G.A.T., & Hendrawati, L.Y. 2015. *Analisis Resiko Kanker Payudara Bersadar Riwayat Pemakaian Kontrasepsi Hormonal dan Usia Jurnal Berkala Epidemiologi, Vol. 3, No. 1 Januari 2015: 12–23, 3, 12–23.*
- Dixit, A., & Hegde, N. P. 2013. *Image texture analysis. International Conference on Advanced Computing and Communication Technologies, ACCT, 69–76.*
- Ermawati, E. 2020. *Klasifikasi Nodul Payudara Berdasarkan Ciri Tekstur pada Citra Ultrasonografi Menggunakan Scilab*. Skripsi. Semarang: Fakultas SAINTEK Universitas Islam Negeri Walisongo Semarang
- Evelyn C. Pearce. 2011. *Anatomi dan Fisiologi Untuk Paramedis*. Jakarta: PT Gramedia Pustaka Utama.

- Fadlli, A., & Rosadi, M. I. 2018. *Klasifikasi Penyakit Jantung Koroner Menggunakan Seleksi Fitur dan Support Vector Machine*. *Jurnal Explore IT* vol.10
- Frank, E., Hall, M., Holmes, G., Kirkby, R., & Witten, I. H. 2010. *Data Mining and Knowledge Discovery Handbook*. ResearchGate 1-10
- Fajrin, H.R. 2016. *Perbandingan Metode Untuk Perbaikan Kualitas Citra Mammogram*. *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*
- Globocan *Breast Cancer Estimated Incidence, Mortality, and Prevalence Worldwide in 2020*. <https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf> diakses 1 April 2021
- Gunawan, Andrey. 2021. *Pemilihan Pemeriksaan Imaging untuk Skrining Karsinoma Mammae*. *Cermin Dunia Kedokteran* vol 48
- Haralick, R. M., Dinstein, I., & Shanmugam, K. 1973. *Textural Features for Image Classification*. *IEEE Transactions on Systems, Man and Cybernetics, SMC-3* (6), 610–621.
- Hogg, P., Kelly, J. & Mercer, C. 2015. *Digital mammography: A holistic approach*. Diunduh di <https://doi.org/10.1007/978-3-319-04831-4> tanggal 15 Maret 2021.
- Houcque,D. 2005. *Introducing To Matlab for Engineering Students*. Northwestern : Northwestern University

- Husnul Khatimi, & Yuslena Sari. 2016. *Prediksi Kualitas Hasil Hut-an Lahan Basah Menggunakan Backpropagation*. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)* 34-42
- Indrati, A., Madenda, S., & Missaoui, R. 2009. *Representasi bentuk tumor payudara dengan kode rantai*
- Kadir, A., & Susanto, A. 2013. *Teori dan Aplikasi Pengolahan Citra*. Yogyakarta: ANDI Yogyakarta
- Kementerian Kesehatan Republik Indonesia. 2019. *Panduan Penatalaksanaan Kanker Payudara (Breast Cancer Treatment Guideline)*. *Jurnal Kesehatan Masyarakat*, 4(4), 1-50
- Laksono, P.J.T. 2017. *Penerapan Forward Selection Pada Support Vector Machine Untuk Klasifikasi Kanker Payudara*
- Listyaningrum, R., Santoso, I., & Isnanto, R. R. 2010. *Analisis Teks-tur Menggunakan Metode Transformasi Paket Wavelet*
- ETP, Lussiana dkk. 2013. *Klasifikasi Citra Mammogram berbasis Teks-tur Menggunakan SVM* *Journal Forum Informatika Kesehatan Indonesia (FIKI 2013)*
- Marques P. 2018. *Medical Image Analysis and Informatics Computer Aided Diagnosis and Therapy*. New York : CRC Press
- Mulyani. 2013. *Kanker Payudara dan Pencegahan Kanker Payudara*. Yogyakarta: Graha Ilmu
- Novar Setiawan, K., & Suwija Putra, I. M. 2018. *Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Sup-*

*port Vector Machine (SVM). Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi). 6(1): 13*

Nugroho, A. 2015. *Klasifikasi Nodul Tiroid Berbasis Ciri Tekstur pada Citra Ultrasonografi*. Tesis. Yogyakarta: Fakultas Teknik Universitas Gadjah Mada

Öztürk, Ş., & Akdemir, B. 2018. *Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA. Procedia Computer Science, 40-46*

Panesar, A. 2021. *Machine Learning and AI for Healthcare*. Conventry : Apress.

Putra, D. K., Santoso, I., & Zahra, A. A. 2009. *Keberadaan Kanker Pada Citra Mammografi Menggunakan Metode Wavelet Haar Transmisi, 11(2), 100-106-106*

Riyanto, Sigit. dkk. 2020. *Pengolahan Citra Medis*. Yogyakarta : CV BUDI UTAMA.

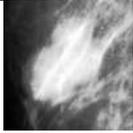
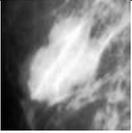
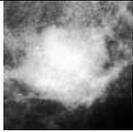
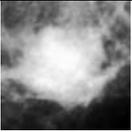
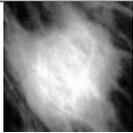
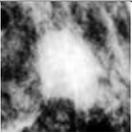
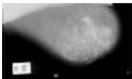
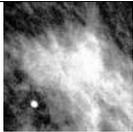
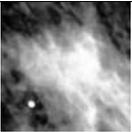
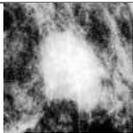
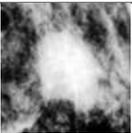
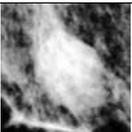
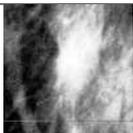
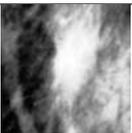
Rosadi, M.I. 2016. *Klasifikasi Massa Pada Citra Mammogram Menggunakan Kombinasi Seleksi Fitur F-Score Dan Ls-Svm*. In *Teknologi (Vol. 6, Issue 1)*.

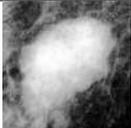
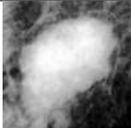
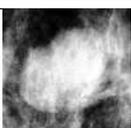
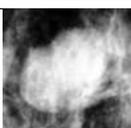
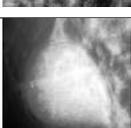
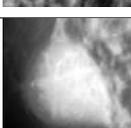
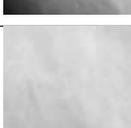
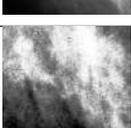
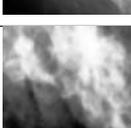
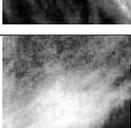
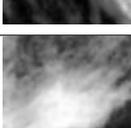
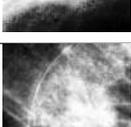
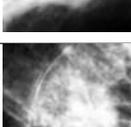
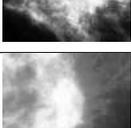
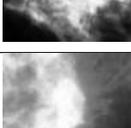
Santoso, B. 2014. *Perancangan Pesawat Sinar-X Mamografi Digital*

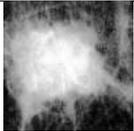
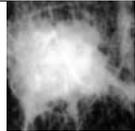
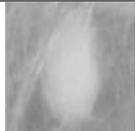
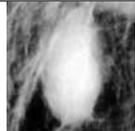
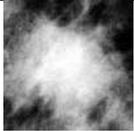
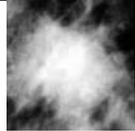
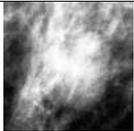
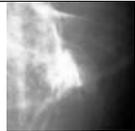
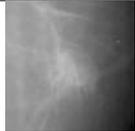
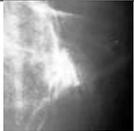
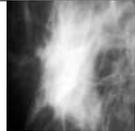
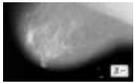
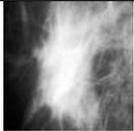
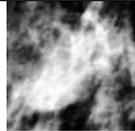
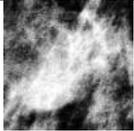
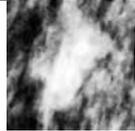
Sari, Dewi Purnama., & Gumayesty, Yeyen. 2016. *Faktor-Faktor Yang Berhubungan Dengan Kejadian Kanker Payudara di Poliklinik Onkologi RSUD Arifin Achmad Provinsi Riau*. Al-Tamimi Kesmas Volume 05

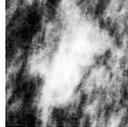
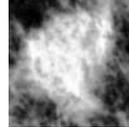
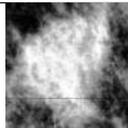
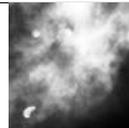
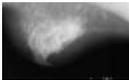
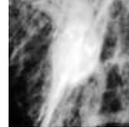
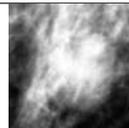
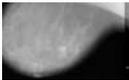
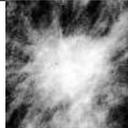
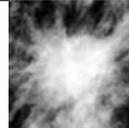
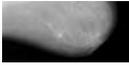
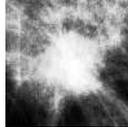
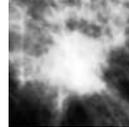
- Setyorini, M. A. E. 2020. *Analisis Perbandingan Metode Machine Learning : Random Forest dan Support Vector Machine untuk Deteksi Kanker Paru-paru*. Skripsi. Jember : FMIPA Universitas Jember
- Suyanto, D., Ramadhani, K. N., & Mandala, S. 2019. *Deep Learning Modernisasi Machine Learning untuk Big Data*. Bandung : Informatika Bandung
- Verma, V., Muttoo, S. K., & Singh, V. B. 2020. *Multiclass malware classification via first- and second-order texture statistics*. *Computers and Security*, 97
- Mao, Youxin. 2010. *Biomedical Imaging*. Croatia : InTech
- Yusufiyah, H.K.N. 2016. *Klasifikasi nodul payudara berbasis ciri bentuk pada citra ultrasonografi*. Tesis. Yogyakarta: Fakultas Teknik Universitas Gadjah Mada
- Zhang, W., & Chen, J. 2018. *Relief Feature Selection and Parameter Optimization for Support Vector Machine based on Mixed Kernel Function*. *International Journal of Performability Engineering*. vol. 14

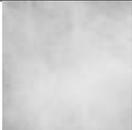
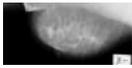
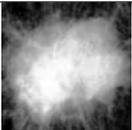
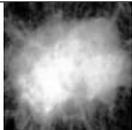
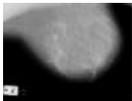
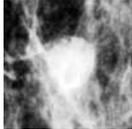
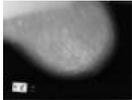
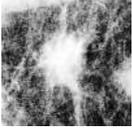
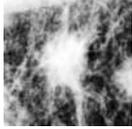
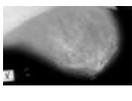
### Lampiran 1. Citra Hasil Pengolahan

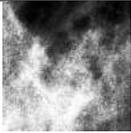
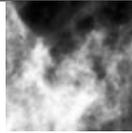
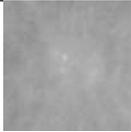
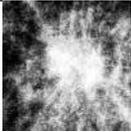
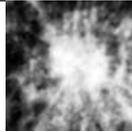
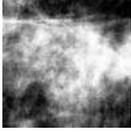
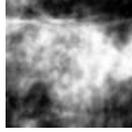
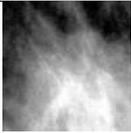
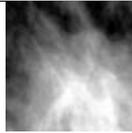
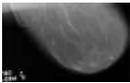
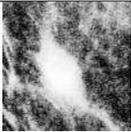
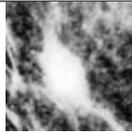
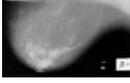
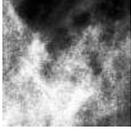
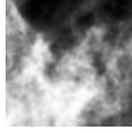
Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terkla-sifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

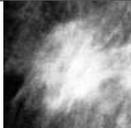
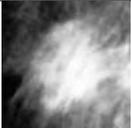
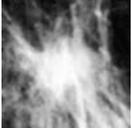
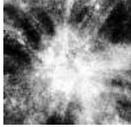
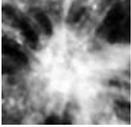
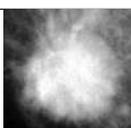
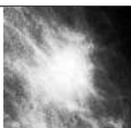
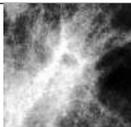
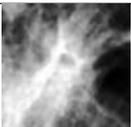
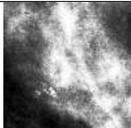
Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>

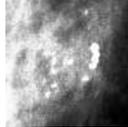
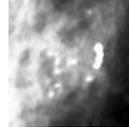
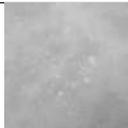
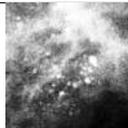
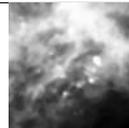
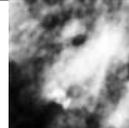
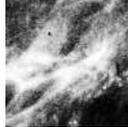
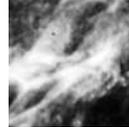
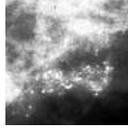
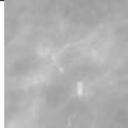
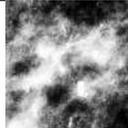
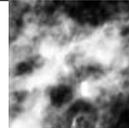
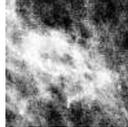
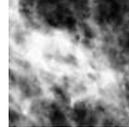
Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>malignant</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

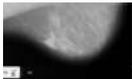
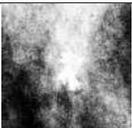
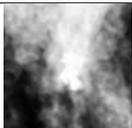
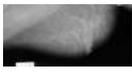
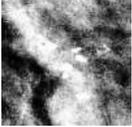
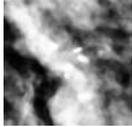
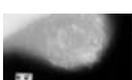
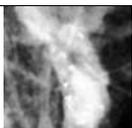
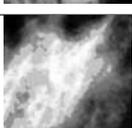
Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>

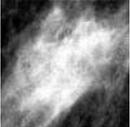
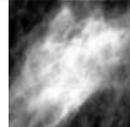
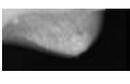
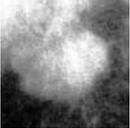
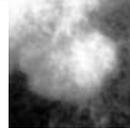
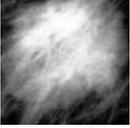
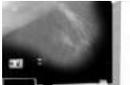
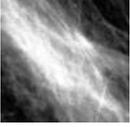
Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>malignant</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>
				<i>benigna</i>	<i>benigna</i>

Citra awal	Citra Crop	Ekualisasi Histogram	Median Adaptif Filter	Kategori	Terklasifikasi
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>
				<i>malignant</i>	<i>benigna</i>





























## Lampiran 3. Hasil Seleksi Ciri

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Attribute Evaluator

Choose ReliefFAttributeEval - M-1-D-1-K-10

Search Method

Choose Ranker -T-1.7978831348823157E308-N-1

Attribute Selection Mode

Use full training set  
 Cross-validation Folds: 10  
 Seed: 1

No class

Start Stop

Result list (right-click for options)

Attribute selection output

```

Instances: 67
Attributes: 150
               [list of attributes omitted]
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 150 GAMBAR/VELAS):
  Relief Ranking Filter
  Instances sampled: all
  Number of nearest neighbours (k): 10
  Equal influence nearest neighbours

Ranked attributes:
0.01851315555639675    102 IDM135
0.01851315555639675    54  IDM45
0.01851315555639675    36  IDM0
0.01851315555639675    78  IDM60
0.018508943050851324    77  IDM90
0.018508943050851324    53  IDM45
  
```

Status

NO	CIRI	RANKED
1	IDM135	0.018513156
2	IDM45	0.018513156
3	IDM0	0.018513156
4	IDM90	0.018513156
5	INN90	0.018308943
6	INN45	0.018308943
7	INN135	0.018308943
8	IDMN135	0.018308937
9	IDMN90	0.018308937
10	IDMN0	0.018308937
11	IDMN45	0.018308937
12	INV90	0.018308935
13	INV0	0.018308935
14	INV135	0.018308935
15	INV45	0.018308935
16	INN0	0.018308883
17	RLN0	0.018270293
18	SRE0	0.017600626
19	RLN135	0.016684966
20	SRE135	0.016569042
21	SRHGE0	0.015268781
22	SRHGE90	0.015228096
23	RP0	0.014485476
24	IMCORR 1 0	0.013908225
25	IMCORR 1 90	0.013908225
26	IMCORR 1 45	0.013908225
27	IMCORR 1 135	0.013908225
28	RP135	0.013658388
29	CSHADE45	0.011996424
30	CSHADE0	0.011996424
31	CSHADE90	0.011996424
32	CSHADE135	0.011996424
33	DISSI135	0.011866201
34	DISSI90	0.011866201
35	DISSI45	0.011866201
36	DISSI0	0.011866201
37	DENTRO135	0.011797872
38	DENTRO0	0.011797872

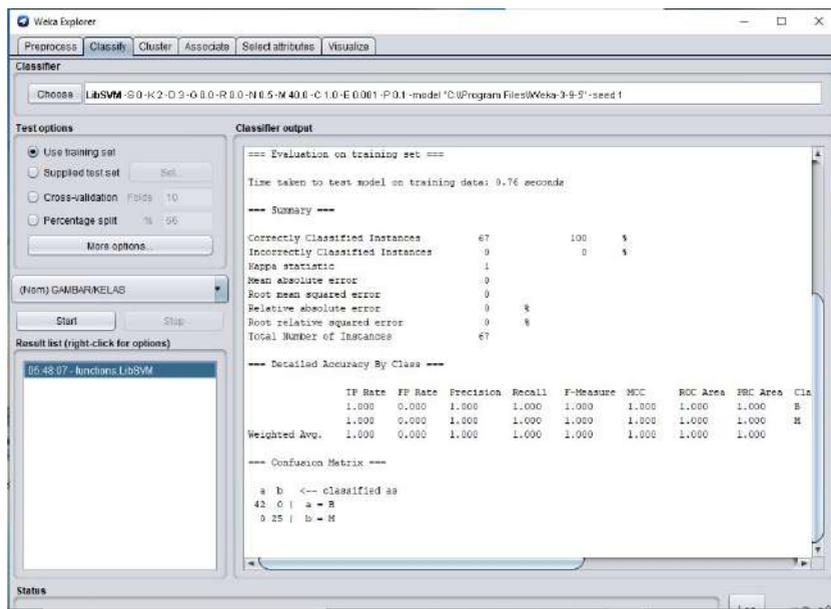
NO	CIRI	RANKED
39	DENTRO90	0.011797872
40	DENTRO45	0.011797872
41	LRE0	0.010583921
42	SRLGE0	0.010000631
43	IMCORR2 45	0.009957486
44	IMCORR 2 135	0.009957486
45	IMCORR 2 0	0.009957486
46	IMCORR 2 90	0.009957486
47	SRLGE45	0.009552353
48	HRGE90	0.009038468
49	SRHGE135	0.0085542
50	HGRE0	0.008400719
51	SAV135	0.008295642
52	SAV0	0.00829564
53	SAV45	0.00829564
54	SAV90	0.00829564
55	CONTRAST135	0.007994903
56	INERTIA135	0.007994903
57	CONTRAST45	0.007994903
58	INERTIA90	0.007994903
59	INERTIA45	0.007994903
60	CONTRAST0	0.007994903
61	CONTRAST90	0.007994903
62	INERTIA0	0.007994903
63	DVAR45	0.007993524
64	DVAR90	0.007993524
65	DVAR0	0.007993524
66	DVAR135	0.007993524
67	JA45	0.007872907
68	JA90	0.007872907
69	JA0	0.007872907
70	JA135	0.007872907
71	JE90	0.0075143
72	JE0	0.0075143
73	JE45	0.0075143
74	JE135	0.0075143
75	LRE135	0.006798039
76	GLN0	0.006167895

NO	CIRI	RANKED
77	HGRE45	0.005954295
78	SRHGE45	0.00567715
79	HGRE135	0.005477528
80	GLN135	0.005009419
81	GLN45	0.004929779
82	GLN90	0.004649892
83	RLN90	0.00455636
84	SRLGE90	0.00452221
85	VARIANCE45	0.004353728
86	VARIANCE0	0.004353728
87	VARIANCE135	0.004353728
88	VARIANCE90	0.004353728
89	SRE90	0.004317256
90	ENTROPI	0.004058992
91	LRE90	0.003909169
92	JV45	0.003777645
93	JV135	0.003777645
94	JV90	0.003777645
95	JV0	0.003777645
96	RP90	0.003698817
97	LRHGE0	0.003640357
98	ACORR0	0.003563289
99	ACORR45	0.003563289
100	ACORR90	0.003563289
101	ACORR135	0.003563289
102	SMOOTHNESS	0.003451269
103	CORRELATION135	0.003438664
104	CORRELATION90	0.003438664
105	CORRELATION0	0.003438664
106	CORRELATION45	0.003438664
107	VARIANS	0.003364438
108	LRHGE135	0.003113957
109	SVAR135	0.002791516
110	SVAR90	0.002791516
111	SVAR45	0.002791516
112	SVAR0	0.002791516
113	SENTRO135	0.002216538
114	SENTRO0	0.002216538

NO	CIRI	RANKED
115	SENTRO90	0.002216538
116	SENTRO45	0.002216538
117	ASM0	0.001806018
118	ASM135	0.001806018
119	ASM45	0.001806018
120	ASM90	0.001806018
121	LGRE0	0.001158432
122	LRLGE90	0.00095811
123	LRHGE90	0.000752441
124	SKEWNESS	0.000258647
125	MPROB135	0.00020872
126	MPROB90	0.00020872
127	MPROB45	0.00020872
128	MPROB0	0.00020872
129	ENERGI	0
130	MEAN	0
131	CPROM45	-4.14E-19
132	CPROM0	-4.14E-19
133	CPROM90	-4.14E-19
134	CPROM135	-4.14E-19
135	SRE45	-0.00080803
136	LRLGE135	-0.000842535
137	KURTOSIS	-0.000940509
138	MODUS	-0.001088162
139	RLN45	-0.0012922
140	LRLGE45	-0.002021939
141	RP45	-0.002300092
142	LGRE45	-0.002895544
143	LGRE135	-0.00389716
144	LRE45	-0.004122051
145	LGRE90	-0.004228638
146	SRLGE135	-0.004808617
147	LRHGE45	-0.006570526
148	LRLGE0	-0.007379009
149	STD	-0.008090372



## Lampiran 4. Hasil Klasifikasi pada *machine learning* Weka



**Classifier**

Choose: **LibSVM-S0-K2-D3-G0-R0-N0.5-N400-C1.0-E0.001-P0.1-model^C:\Program Files\Weka-3.9.5\seed1**

**Test options**

- Use training set
- Supplied test set
- Cross-validation Folds: 10
- Percentage split %: 56

(Nom) GAMBAR KELAS:

**Result list (right-click for options)**

06.48.07 - functions.LibSVM

**Classifier output**

```

=== Evaluation on training set ===
Time taken to test model on training data: 0.76 seconds

=== Summary ===
Correctly Classified Instances      67      100 %
Incorrectly Classified Instances    0         0 %
Kappa statistic                     1
Mean absolute error                  0
Root mean squared error              0
Relative absolute error              0 %
Root relative squared error          0 %
Total Number of Instances          67

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  B
1.000  0.000  1.000  1.000  1.000  1.000  1.000  1.000  M
Weighted Avg.  1.000  0.000  1.000  1.000  1.000  1.000  1.000

=== Confusion Matrix ===
 a b  <-- classified as
42  0 | a = B
 0 25 | b = M
  
```

**Status**

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: **LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-9-5\seed1**

Test options

Use training set

Supplied test set

Cross-validation Folds:

Percentage split %:

(Nem) GAMBARKELAS

Result list (right-click for options)

05:48:07 - functions.LibSVM

05:52:04 - functions.LibSVM

Status

Classifier output

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 42 62.6966 %

Incorrectly Classified Instances 25 37.3134 %

Kappa statistic 0

Mean absolute error 0.3731

Root mean squared error 0.6108

Relative absolute error 79.4561 %

Root relative squared error 126.0705 %

Total Number of Instances 67

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Clas
	1.000	1.000	0.627	1.000	0.771	?	0.500	0.627	B
	0.000	0.000	?	0.000	?	?	0.500	0.373	M
Weighted Avg.	0.627	0.627	?	0.627	?	?	0.500	0.532	

--- Confusion Matrix ---

a b <-- classified as

42 0 | a - B

25 0 | D - M

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: **LibSVM -S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1 -model "C:\Program Files\Weka-3-9-5\seed1**

Test options

Use training set

Supplied test set

Cross-validation Folds:

Percentage split %:

(Nem) GAMBARKELAS

Result list (right-click for options)

05:48:07 - functions.LibSVM

05:52:04 - functions.LibSVM

05:54:31 - functions.LibSVM

Status

Classifier output

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 42 62.6966 %

Incorrectly Classified Instances 25 37.3134 %

Kappa statistic 0

Mean absolute error 0.3731

Root mean squared error 0.6108

Relative absolute error 79.4561 %

Root relative squared error 126.0705 %

Total Number of Instances 67

--- Detailed Accuracy By Class ---

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Clas
	1.000	1.000	0.627	1.000	0.771	?	0.500	0.627	B
	0.000	0.000	?	0.000	?	?	0.500	0.373	M
Weighted Avg.	0.627	0.627	?	0.627	?	?	0.500	0.532	

--- Confusion Matrix ---

a b <-- classified as

42 0 | a - B

25 0 | D - M

## Lampiran 5. Source Code Scilab

### 1. Ekualisasi Histogram

```
clc;
clear;
a=uigetfile('.jpg','Select the Image:-');
b=imread(a);
x=imresize(b,[256,256]);
j=imhistequal(b);
//figure(); imshow(j);
imwrite(j,"D:
SKRIPSI
proses
hisequ
M25.jpg")
```

### 2. Median Adaptif Filter

```
clc;
clear;
a=uigetfile('.jpg','Select the Image:-');
H=imread(a);
//b=imresize(b,[256,256]);
//H=rgb2gray(c);
x=immedian(H,[9,9]);
xa=immedian(H,[9,9]);
[baris, kolom]=size(H);
t=baris*kolom
```

```

m=0;
Tl=m/t
for bar=1:baris
for kol=1:kolom
intensity=H(bar,kol);
if intensity <=Tl
xs(bar,kol)=xa(bar,kol); else
xs(bar,kol)=x(bar,kol);
m=m+H(bar,kol);
end
end
end
figure; imshow(xs);
imwrite(xs,"D:
SKRIPSI
proses
median adaptif
M25.jpg")

```

### 3. Histogram

```

clc;
clear;
D=uigetfile(".jpg",
"Select The Image");
F=imread(D);
[m, n]=size(F);

L=256;

```

```

Frek=zeros(L,1);
//Menghitung frekuensi aras keabuan F=double(F);
for i=1:m
for j=1:n
intensitas=F(i,j);
Frek(intensitas+1)=Frek(intensitas+1)+1;
end
end

//Menghitung probabilitas  $jum_{piksel}=m^{*}n$ ;
for i=0:255
 $Prob(i + 1) = Frek(i + 1)/jum_{piksel}$ ;
end
//disp(Prob(i + 1));

```

#### 4. GLCM

```

clear;// clc;//
g=uigetfile('.jpg','Select the Image:-');
gam=imread(g);
ga=imresize(gam,[256, 256]);
//gr = rgb2gray(ga);
 $F = int16(ga)$ ;

// $F = [0,0,,1,1;0,0,1,1;0,2,2,2;2,2,3,3]$ ;
//disp(F);
[tinggi,lebar]= size(F);

//bentukGLCM

```

```

GLCM135 = zeros(tinggi, lebar);
totalpiksel135 = 0;
for y=1:tinggi
    for x=1:lebar
        //sudut135
        if (x<lebar) then
            a = F(y,x); a = iconvert(a, 12);
            b = F(y,x+1); b = iconvert(b, 12);
            printf("y=GLCM135(a+1, b+1)=GLCM135(a+1, b+1) + 1;
            totalpiksel135=totalpiksel135 + 1;
        end
    end
end
GLCM135T=GLCM135';
G135=GLCM135 + GLCM135T;
jumlah135=sum(G135);
HGLCM135=G135/jumlah135;
//disp(HGLCM135);

```

## 5. GLRLM

```

clear; //clc; //
g=uigetfile('.jpg','Select the Image:');
gam=imread(g);
ga=imresize(gam,[256, 256]);
//gr=rgb2gray(ga);
F=int16(ga);

```

```

[Nx, Ny]=size(F);
L1=unique(F)';
nL1=size(L1,"*");

maxrunlength135=Nx
GLRLM135=zeros(256,maxrunlength135)

for m=1:Nx
for n=1:Ny
graylevel=F(i,j)
for l=1:nL1
if L1(l)==graylevel then
NoGrayLevel=l
end
end
if (i==Nx) || (j==Ny) || (i<Nx & j<Ny & graylevel<>F(i+1,j+1))
then
for k=0:j-1
else
break
end
end
end
GLRLM135(NoGrayLevel,runlength)=GLRLM135(NoGrayLevel,runlength)
end
end
end
//disp(F)
//disp(GLRLM135)

```



## Lampiran 6. Riwayat Hidup

### A. Identitas Diri

1. Nama : Uhty Maesyaroh
2. Tempat, Tanggal Lahir : Grobogan, 21 Mei 1998
3. Alamat : Desa Tlaga 01/05 Gumelar, Banyumas, Jawa Tengah
4. No. Hp : 0813-1671-0668
5. Email : uhtymay@gmail.com

### B. Riwayat Pendidikan

#### 1) Pendidikan Formal

- a. MI Nur Attaqwa (2003-2009)
- b. SMPN 114 Jakarta (2009-2012)
- c. SMAN 75 Jakarta (2012-2015)

#### 2) Pengalaman dan Organisasi

- a. Panitia Olimpiade Fisika Islam Nasional (OFIN) UIN Walisongo Semarang (2016 dan 2017)
- b. Ikatan Mahasiswa Fisika UIN Walisongo Semarang (2017-2018)
- c. Keluarga Mahasiswa Banyumas (2018-2019)

Semarang, 1 Juli 2022



Uhty Maesyaroh

NIM : 1608026010